ABSTRACT

Title of Document:

Distribution and Evolution of the Palatinose (*pal*) operon in *Enterobacter sakazakii*

Katelyn Amber Lamm,
Master's in Food Science
2008

Directed By: Professor Jianghong Meng
Department of Nutrition and Food Science

Enterobacter sakazakii, an opportunistic enteric pathogen found in infant milk formula, is associated with outbreaks of meningitis and septicemia among immunocompromised newborns. A biochemical signature of *E. sakazakii* metabolism is its ability to degrade palatinose. The prevalence of the *pal* operon among *E. sakazakii* remains unknown. Studying 144 *E. sakazakii* strains, the prevalence of five *pal* genes (*palZ*, *K*, *Q*, *E*, and *G*) were assessed. Only 22% of the strains retained all five genes in the *pal* operon, while another 22% of strains retained four of the five genes. Genetic instability of the *pal* operon was seen in phylogenetic studies which revealed evidence for horizontal gene transfer of the *pal* genes among strains and revealed variation in the structure and distribution of the *pal* operon. The genetic instability of the *pal* genes inferred from this study rejects using this operon as a diagnostic target for *E. sakazakii* detection in infant formula.

Distribution and Evolution of the Palatinose (*pal*) operon in *Enterobacter sakazakii*

By

Katelyn Amber Lamm

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Masters of Food Science
2008

Advisory Committee:
Professor Dr. Jianghong Meng, Chair
Dr. Eric W. Brown
Associate Professor Dr. Y. Martin Lo

Preface

Enterobacter sakazakii is an emerging food borne pathogen in the infant formula industry. While its incidence for causing disease remains rare, the illness has cases cause serious health implications. Due to increasing antibiotic resistance and the seriousness of the *E. sakazakii* infection in infants, prevention remains our most effective weapon against this pathogen. Since the palatinose operon is being used as a diagnostic marker to identify the organism, its prevalence and stability in the *E. sakazakii* family warrants research attention, which led to this study.

Foreword

'Human intelligence, culture and technology have left all other plant and animal species out of the competition. We also may legislate human behavior. But we have too many illusions that we can, by writ, govern the remaining vital kingdoms, the microbes, that remain our competitors of last resort for dominion of the planet. The bacteria and viruses know nothing of national sovereignties. In that natural evolutionary competition, there is no guarantee that we will find ourselves the survivor'.

----Joshua Lederberg, Nobel Laureate

Acknowledgements

The author wishes to express her thanks to the University of Maryland Department of Nutrition and Food Science for the opportunity to receive a Master's Degree in Food Science. She would like to thank Dr. Jianghong Meng for advising her and extending to her the opportunity to conduct her research at the FDA's Center for Food Safety and Applied Nutrition.

The author would like to thank Dr. Martin Lo for serving on her committee and for his help in many areas during her time at this University. His advice and guidance were greatly appreciated.

The author would like to extend her deep gratitude to Dr. Eric Brown for his tireless efforts in guiding her research, his real world advice, his moral support, his humor, and his making her apart of his FDA family. His spirit and drive for scientific research is appreciated as much as his salacious vocabulary. Dr. Brown is a priceless, aberrant, one of a kind man of science.

The author would like to thank Mr. Steven Miller for being the foundation for her love of science and for instilling in her the drive to pursue a career in this field. She would also like to thank him for going above and beyond the work of a teacher and instilling in me the real world knowledge to make it. From the first E. coli streak to the first gram strain to the first 'yehoo' which led me to this day, the author thanks him.

The author would also like to thank Mr. Raymond Edwards for being the first to expose her to the world of genetics and food science, both of which have led her to

this thesis research. She would also like to thank him for proper instruction on writing a summary. She keeps the mini skirt in mind to this day.

The author would like to thank Penn State University for giving her the opportunity to attend Governor's School and to achieve a Bachelor's Degree in Food Science. The memories are endless.

The author would like to thank all of her Penn State friends for getting her through school. She would like to thank Sarah Erdman for being her constant companion through labs and through Rome; Sarah Augustino for being an endless study companion at the HUB; Emma Cooper for Rome and studying with a white board; Megan Oberdick for her comic relief and nutrition crisises; to Jeffrey Lombard for answering my Chem Lab questions, for teaching her organic chemistry, for instilling in her a fear of Siamese cats and for being the constant support to get her through. L'chayim! To all of the nittany lions that made the writer's Penn State experience priceless, there isn't enough space to thank you enough.

The author would like to thank her family for being there from the beginning of it all. She would like to thank her mother for instilling in her a love of agriculture and for never letting her settle for anything less then her best. She would like to thank her father for his endless curiosity and his small ways of always showing his daughter he is there. She thanks both her parents for never failing and always putting their daughter first. The author knows she will never be able to repay them but she hopes to try. The author would like to thank her sister for loving her for who she is and for being an ever-present beacon of support and love.

The author would like to thank her grandmother for her realism, honesty, and frankness. She would also like to thank her for showing the writer what she will inevitably be like in 50 years. The author would like to thank her grandfather for being the sweetest, kindest, most patient man she has ever met. She would like to thank him for teaching her to sing, to garden, to play cards, to spell, and to always look for pennies.

The author would like to thank Drew Brown for his love, patience, support, and laughter. She would like to thank him for his everlasting support and for being able to cook. Thank you for being my best friend.

Table of Contents

List of Tables

List of Figures

Figure 9: The phylogenetic tree of *palE* in *E. sakazakii*. The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279277).

Figure 10: Phylogenetic Tree of *palQ* gene in *E. sakazakii*. The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279283).

Figure 11: *E. sakazakii* MLST tree of the 5 housekeeping genes. (*icdZ*, *aspC*, *gyrA*, *lysP, clpX*). The MLST tree reflects the overall taxonomy of the *E. sakazakii* strains studied here. The tree shown reflects the minimum number of gains and losses of the 5 *pal* genes during the evolution of *E. sakazakii*.

Figure 12: Pairwise comparison tree of *palZ* and *palK*. Strain 1514 and 1673 were identified through ILD tests to be causing incongruence. Strains 1680, 1563, and 1676 were identified by sight as being probable causes of incongruence. It can be seen that they are in different clades in the opposite trees, indicating horizontal gene transfer, while the other strains are located in virtually the same spots on both trees.

Figure 13: Pairwise comparison of *palZ* and *palQ* trees. Strains 1681, 1680, and 1563 were identified as incongruent by the ILD test. 1672 was identified by sight as being in a different clade in *palZ* as opposed to *palQ*.

Figure 14: Pairwise comparison of *palQ* and *palK* trees. Strain 1672 was identified by the ILD test. Strains 1677, 1563 and 1553 were identified by sight as causing incongruence.

Figure 15: Pairwise comparison of *palG* and *palZ* trees. Strain 1563 was identified as causing incongruence by the ILD test, where as strain 1528 was identified by sight as causing incongruence.

Chapter 1: Introduction

Enterobacter sakazakii is an emerging foodborne pathogen that cause illness in infants mainly through dried infant milk formula. This organism can cause various diseases such as meningitis, bacteremia, and urinary tract infections in infants, especially those which are underweight or premature. The endotoxin of *E. sakazakii* can cause endotoxemia, and the pathogen can also cross the blood brain barrier, making its infections more severe with lasting complications.

The environmental reservoir of *E. sakazakii* is still largely unknown, as are many other characteristics of the organism. A reliable detection method is currently being perfected to ensure the safety of dried infant formula through early detection of the pathogen.

One characteristic that separates *E. sakazakii* from other species in the *Enterobacter* family is its ability to metabolize palatinose using the α-glucosidase operon, which is a cluster of at least five genes. Palatinose is a sucrose isomerase which cannot be metabolized by most organisms, and is therefore advantageous as a way to take a carbon source away from other organisms. It has also been found that in *Erwinia rhapontici*, a near neighbor which contains these genes, the *pal* operon is activated by the presence of palatinose but inhibited by sucrose presence (Bornke 2001).

Many diagnostic methods for *E. sakazakii* are based on the presence of α -glucosidase activity. However, the operon has been found in some related enteric species and conversely has not been found in some *E. sakazakii* strains. Taken together, these problems can lead to many false positives and false negatives.

The purpose of this study was first and foremost to learn about the distribution and patterns of acquisition of the α-glucosidase operon in *E. sakazakii* and related strains. Since preliminary data have shown that some strains only have a few genes of the operon, the effects of this on the phenotypic display of the α-glucosidase activity was also be examined. A phenotypic microarray was be used to determine what combination of the genes led to the expression of the phenotype to be able to metabolize palatinose, and what combination did not.

Chapter 2: Literature Review

Background

Enterobacter sakazakii is an emerging foodborne pathogen that infects infants through dried infant milk formula (Kang 2008, Derzelle 2007). It is a Gram-negative rod that produces an endotoxin and invades the infant gut, leading to meningitis, bacteremia, and urinary tract infections (Iverson 2007). *E. sakazakii* infections have a mortality rate of 40-80% (Corti 2007). It has also been associated with necrotizing enterocolitis, but has not been found to be the causative agent (Ray 2007). Infants are most at risk if they are less than 28 days old and if they are of low birth weight, which is classified as less than 2500 g (Mullane 2007). While infections in adults are rare, they have been reported, and usually in the elderly (See 2007).

E. sakazakii was originally referred to as the yellow pigmented *Enterobacter cloacae*, but it was named *Enterobacter sakazakii* in 1980 after more genetic and biochemical differences were discovered between the two organisms. Its environmental source is unknown, which poses a problem in preventing contamination (Seo 2005 and Gurtler 2005). It has been isolated from a variety of food, floors and drains of manufacturing plants, in the air and equipment in hospitals, rats and soils (Shaker 2006, Lin 2007). In another study, it was isolated from plant foods, cereal, fruit, vegetables, legumes, herbs and spices (Friedemann 2007).

While *E. sakazakii* is known for contaminating infant formula, studies have also found the pathogen in dried fish products. Dried shrimp were found to have the worst contamination (Kim K 2008).

Several of the factors that distinguish *E. sakazakii* from other *Enterobacter* species include its inability to ferment sorbitol, the production of yellow-pigmented colonies on trypticase soy agar and the activity of α -glucosidase (Lehner 2006).

Association with Infant Formula

E. sakazakii has been found to be associated with dried infant formula milk. In a study in which 75 IFM products were sampled, the majority of the products had less than 10^3 cfu/g and 1% of the samples had greater than 10^4 cfu/g of *E. sakazakii*. *E. sakazakii* could be recovered from infant formula, the environment it was prepared in, and in unopened infant formula products (Baiguini 2005). It has been found to be a post pasteurization problem, with contamination occurring after heat treatment (Drudy 2006).

A study was conducted to examine the frequency of *E. sakazakii* in infant formula manufacturing facilities. Some of the specific locations in the plant had a 31% contamination rate. The overall frequency of intermediate and final infant formula product was 2.5% (Mullane 2007). Another study was conducted to examine the prevalence of *E. sakazakii* in various formulas. One of the infant formula types had a contamination rate of 92% (Torres-Chavolla 2007).

One factor that enables *E. sakazakii* to contaminate dried infant formula is its tolerance to dry stress. A study was done to compare dry stress survival time in various bacterial species. *Citrobacter. freundii*, *Citrobacter koseri*, and *Enterobacter cloacae* could survive no longer then 6 months in dried infant formula. *Salmonella* Enterica, *Klebsiella. pneumonia*, and *Escherichia coli* could survive up to 15 months.

Pantoae spp., *Klebsiella. oxytoca*, and *Escherichia. vulneris* could survive up to 15 months. However, *E. sakazakii* could be found in infant formula after 2.5 years (Barron 2007).

Another study was done to test the survival and recovery of *E. sakazakii* in dried infant formula after 3-10 days. Lag times were found to be short. At 10°C, the lag time was found to be 83.3 ± 18.7 hours, and at 37°C the lag time was found to be 1.73 ± 0.43 hours (Kendhai 2006).

E. sakazakii has also been found to be resistant to osmotic pressure. Desiccated cells have an accumulation of protein that most likely perform some kind of protective function. Cells under osmotic pressure were also found to down regulate motility apparatus and flagella (Reidel 2007).

Heat resistance has been associated with the pathogen having a higher level of translation initiation factor, which is coded by gene infB. The especially heat resistance strains had higher amounts of this than the heat sensitive strains (Asakura 2007).

The lethal dose of the organism is another factor that is not known. In a study done with mice, the lethal dose when the organism was taken orally was 10^7 cfu/g. When injected, the lethal dose was 10^5 cfu/g (Gurtler 2005). Symptoms of the toxicoinfection include meningitis, brain abscesses, bulging fontanelles, seizures, encephalitis, convulsions, instability of body temperature, and hydrocephalus (Gurtler 2005).

Studies have been done to prevent *E. sakazakii* infection when it is present. One method to prevent the growth of the pathogen would be to ensure that there is no

temperature abuse of the formula (Forsythe SJ 2005). Rehydrated infant formula should be held for no longer than 4 hours, and it should be stored at 4°C or less so that *E. sakazakii* cannot grow (Gurtler 2007, Lin 2007). However, it was found that most refrigerators are kept at 10°C. Vegetative cells can also be reduced or eliminated by rehydrating the infant formula with water at 60°C, which is higher then the manufacturer recommendation of 50°C (Kim SH 2007). The lactoperoxidase preservation system was also found to be effective against *E. sakazakii*, however this would be inactivated with pasteurization (Gurtler 2007).

Infection & Role of Endotoxin in Pathogenicity

Of the children infected with *E. sakazakii*, 75% are less than 1 month old and are often infants of low birth weight or that were born premature. The risk for infants being infected with *E. sakazakii* is multifactorial. They have an immature immune system and their gut is not yet colonized with native microflora, leaving pathogens the opportunity to invade and colonize (Townsend 2006). *E. sakazakii* uses actin filaments and microtubule structures to adhere to and invade the Caco-2 epithelial cells (Mange 2006, Kim KP 2008). The *ompA* gene was also discovered as being an important factor in the invasion of INT407 epithelial cells (Mohan 2007).

The endotoxin produced by the lipopolysaccharide system of its gram negative wall has been shown to damage the lining of the colon, promote bacterial invasion and weaken the blood brain barrier leading to other infections and complications, such as meningitis, bacteremia, and urinary tract infections. The endotoxin is heat stable at 100°C and can therefore remain stable when the infant

formula is heat treated to kill vegetative cells (Townsend 2006). In another study, the endotoxin was found to be stable at 90°C for 30 minutes at pH 6 (Raghav 2007).

The endotoxin also causes histological changes in the epithelial lining of the colon, which become thinned and flattened (Townsend 2006). The endotoxin also impairs the migration of enterocytes to the location of the damage in the epithelial lining. In some cases, the pathogen was associated with necrotizing enterocolitis, the inflammation of the colon, which later leads to tissue death. *E. sakazakii* infections have been found to be 20 times more frequent in infants fed formula as opposed to infants fed breast milk (Townsend 2006). Human breast milk can also become contaminated with *E. sakazakii* through breast milk fortifiers (Lenati 2008). *E. sakazakii* growth is not influenced by the composition of the dried formula (Gurtler 2007).

E. sakazakii can cause an infection alone, but the infection can be worsened by its production of endotoxin, which can cause health problems of its own without the pathogenic cells being present. Endotoxemia is the immunological response to the presence of endotoxin in the blood stream. Symptoms of this are fever, low blood pressure, leukocytosis, thrombocytopenia, and coagulopathies (Townsend 2006).

The organism with the endotoxin, just the organism, or just the endotoxin can be found in infant formula, and the endotoxin is not killed with the usual heat treatment administered to dried infant formula. Since the reservoir of the organism is not known, the source of contamination in the formula is also not known. The endotoxin could be inoculated into the formula with the water used to rehydrate it. There is no correlation between samples with high endotoxin levels and bacterial

counts in infant formula. In infant formula tested, the lowest level of concentration found was 40 Endotoxin Units/g, and the highest level was found to be 5.5 x 10^4 EU/g. Half of the samples had concentrations greater than 3000 EU/g (Townsend 2006). Levels in the United Kingdom tap water were found to be as high was 32 EU/ml (Townsend 2006)

Outbreaks

The first two documented cases of *E. sakazakii* infection were in 1958 in England. It was then described as the yellow pigmented *E. cloacae*. Between the 1980s and 1990s 25 cases of sporadic infant deaths were reported in Tennessee, Maryland, Ohio, Massachusetts and Louisiana. Infant formula was confirmed as responsible. In 1995, Massachusetts reported a five- case outbreak with patients ranging from the ages of 3-82. The three oldest patients died. Another outbreak occurred in Belgium in 1998, which involved twelve cases and two deaths. Infant formula was again confirmed as the source. Another outbreak occurred in 2001 in Tennessee, which involved ten cases and one death from contaminated infant formula (Gurlter 2005). An outbreak in France in 1994 affected thirteen infants and resulted in three deaths. There was also an interesting find in that three of the infants were infected with more then one *E. sakazakii* genotype and four of the infants colonized were symptomless (Caubillo- Barron 2007).

Case Study

A case study was done to compile cases of *E. sakazakii* infections and attempt to learn about risk factors for contracting the pathogen. The case definition was met by 46 infants who had a confirmed *E. sakazakii* infection. Out of these 46 infants, 33 had meningitis, 12 had bacteremia, and 1 had a urinary tract infection. While 19 of the 33 infants with meningitis survived, they had lasting health complications including brain abscesses, developmental delays and motor impairment. Only one of the infants with bacteremia died, however this infant also had necrotizing enterocolitis (Bowen 2006).

Out of the 46 infants infected, 21 of them were born prematurely. Feeding practices were described in 26 of the cases, and 24 of them were fed with powdered infant formula. *E. sakazakii* was cultured from the patient and from the formula. Multiple manufacturers were implicated (Bowen 2006).

This case study showed that while *E. sakazakii* infections are rare, they cause severe complications even after the infection is gone. Prevention of the accidental consumption of *E. sakazakii* is very important, especially due to growing antibiotic resistance of the pathogen.

Antibiotic Resistance

The more that can be discovered about *E. sakazakii*, the more measures that can be taken to prevent its presence and its endotoxin's presence in infant formula. Prevention of infections is very important, since treatment is becoming more complicated.

The gold standard treatment for *E. sakazakii* infections was traditionally Gentamicin and Ampicillin. Gentamicin is a broad spectrum antibiotic for gram negative organisms, however the antibiotic does not achieve adequate concentrations in cerebral spinal fluid, making it less effective in patients with meningitis (Nazarowec-White 1999).

Different strains of *E. sakazakii* have been found to have varying susceptibility or resistance to certain antibiotics. *E. sakazakii* is more susceptible to antibiotics than other *Enterobacter* species, however it is very likely that *Enterobacter* can transfer resistance to *E. sakazakii* horizontally (Nazarowec-White 1999).

In a study done in 1999 on the antibiotic resistance of *E. sakazakii*, several strains were tested for their susceptibility to antibiotics. MNW7 was found to be the most resistant strain, and showed to be unaffected by sulphisoxazol, cephalothin, chloramphenical, and ampicillan. MNW1 and MNW6 were found to be resistant to chloramphenical, ampicillan, sulphisoxazole and cephalothin. LCSC3 was found to be resistant to sulphisoxazole and cephalothin. It was found to be susceptible to ampicillan, ceftoxim, chloramphenical, gentamicin, kanamycin, polymixin B, trimethoprim-sulfamethoxazole, tetracycline, and streptomycin (Nazarowec-White 1999).

Biofilm formation has been found to be an environment that favors gene exchange among bacteria, and therefore enabling the spread of antibiotic resistance. *E. sakazakii* has been found to form biofilms on stainless steel and on enternal feeding tubes. Biofilms were not formed at 12°C but they were formed at 25°C (Kim

H 2006). Cellulose production was found to be a component of biofilm formation on M9 minimal media, and 2 operons responsible for creating cellulose were discovered (Grimm 2008).

Optimal therapies for *E. sakazakii* meningitis have not been determined and currently a combination of ampicillan and gentamicin has been used in treatment. In one study, all strains tested were found to be susceptible to at least 13 antibiotics. However, all strains tested were resistant to vancomycin, cephalzolin and cefpodoxime (Pei 2007). Gentamicin is usually used for general *E. sakazakii* infections since is a broad spectrum gram negative antibiotic, however it is ineffective in treating *E. sakazakii* meningitis because it does not achieve adequate concentrations in cerebral spinal fluid (Nazarowec-White 1999). While antibiotic resistance is a growing problem that needs to be studied, prevention of *E. sakazakii* contamination in infant formula through improved detection methods will prevent much of the need for the use of antibiotics in treating infant infections and complications.

Detection

The current detection method used by the Food and Drug Administration takes at least 6 days to completely confirm the identity of *E. sakazakii* in a sample, as can be seen in Table 1. This method is not completely accurate when differentiating between *E. sakazakii* and other closely related *Enterobacter* species. Tryptone Soy Agar (TSA) was used because it was the original media which the new *E. sakazakii* classification was based on. It was noted that certain strains of *E. cloacae* appeared

to form yellow colonies on TSA. After further studies, significant biochemical and genetic differences were found to give *E. sakazakii* its on classification in 1980 (Seo 2005). *E. helvetica* also appears blue-green on DFI, as does *Escherichia vulneris*. However, when glucose was added to DFI, *E. vulneris* appears white (Song 2008).

Table 1: Current FDA Protocol for Identifying *E. sakazakii*

Step	Time
Dilute samples of IF in water	Overnight
Add 10 mL to enrichment broth and incubate	Overnight
Streak onto VRBG agar and incubate	Overnight
Pick presumptive colonies and streak onto TSA	48-72 hours
Select colonies to conduct biochemical confirmation	N/A
Calculate Most Probable Number (MPN)	N/A
	= at least 6 days

A new method has been proposed by the FDA and is to be approved by AOAC, as can be seen in Figure 2. This method offers a more rapid and accurate approach to identifying the presence of *E. sakazakii*. It uses a Real-Time PCR assay, plating on chromogenic agar, and a second API biochemical test to verify the identification which can be done in a day. Allowing the cells time to recover is important since they were in a dry stressed environment. Some cells may give false negatives if plated directly onto Droggan-Forsythe-Iverson (DFI) due to a lack of recovery time (Al-Holy 2008). The chromogenic plating is done on DFI agar, which uses different colony colors to indentify organisms. For example, *Salmonella* appears black, and *E. sakazakii* appears blue- green. All other gram negative organisms

appear some shade of white, and gram positives do not grow due to vancomycin in the media.

DFI contains indolyl substrate 5-bromo-4-chloro-3-indolyl-α,D-glucopyranoside. α-glucosidase hydrolyzes the compound, releasing the compound 5-bromo-4chloro-indolol. 5-bromo-4-chloro-indolol then reacts with oxygen to form bromo-chloro-indigo, which is the pigment that creates the blue- green color around the *E. sakazakii* colonies.

Table 2: Proposed AOAC Method for *E. sakazakii* Testing

Step	Time
25 g sample PIF taken and inoculated into 225 ml of water	6 hours/samples also incubated for 24 hours to resuscitate dry stressed cells.
10 ml aliquot into tubes	N/A
Tubes centrifuged for 5000 rpm	10 minutes
Supernatant decanted	N/A
Pellet resuspended into 100 ml water	N/A
Pellet plated on Chromogenic agar (DFI) and incubate at 37C	Overnight
Confirm with RT-PCR and Rapid ID	N/A
	= 24 hours

Reducing the time required for identification of the pathogen while improving the accuracy and selectivity is very important for quality assurance in industry. It can ensure the safety of infant milk formula and allow for regulatory agencies to trace and prevent outbreaks.

The sensitivity of the Real-Time PCR assay was found to be as low as 100 CFU/ml. *E. helvetica* also shows up as blue-green on the DFI agar however does not

have the MMS operon (the single specific target for RT-PCR). This operon is the second biochemical confirmatory method which is important to distinguish between the two organisms. *E. turicensis* has the MMS operon and is RT-PCR positive, but shows up negative on DFI agar. It is also important to discover if other organisms have the MMS operon, would appear blue-green on the DFI agar, or tests a false-positive for the biochemical method of choice. Since there are some organisms that can be confused with *E. sakazakii* using only RT-PCR or DFI agar, it is important to use the two methods together. Due to the zero tolerance policy for *E. sakazakii* in infant formula, methods still need to be improved to detect less than 100 CFU/ml. Colonies on a DFI plate is shown in Figure 1.

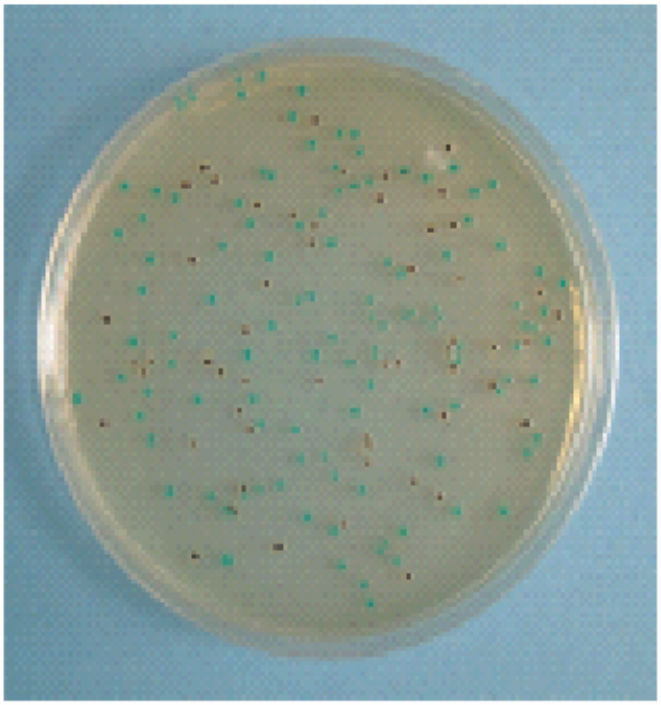


Figure 1: Colonies of *E. sakazakii* and *Salmonella* plated onto DFI agar. *Salmonella* colonies appear black and *E .sakazakii* colonies appear blue-green.

α -*Glucosidase Activity*

α -Glucosidase activity is one of the important biochemical features of *E.sakazakii* that distinguishes it from other species in the *Enterobacteriaceae* family. However, it has been found that a few other organisms also have this activity. α-glucosidase activity has been found to be controlled by a cluster of genes that are responsible for the metabolism of isomaltulose (palatinose; 6-O-α-glucopyranosyl-D-fructose) (Lehner 2006). These genes have also been seen in *E. rhapontici*, and most likely have similar functions as can be seen Table 3. Other than *palK*, *palQ*, *palE*, *palZ* and *palG*, *E. rhapontici* has three extra *pal* genes, including *palI*, *palH*, and *palR*. *palI* functions as another α-1,6-glucosidase. *PalH* has α-galactosidase activity and *palR* is a transcriptional regulator (Lehner 2006).

Table 3: Activity of *pal* genes in *E. rhapontici* and *E. sakazakii* (Lehner 2006).

Gene	Proposed Function
*pal*Z	Family 13 glycosyl hydrolase (exo- α-1,4-glucosidase)
*pal*Q	Family 13 glycosyl hydrolase (oligo-1,6-glucosidase)
*pal*G	In the family of inner membrane permeases
*pal*K	ABC/ATP-binding cassette proteins
*pal*E	In the family of periplasmic sugar binding proteins
*pal*F	In family of inner membrane permeases

α-glucosidase is an amylolytic hydrolase that releases glucose from the non-reducing end of poly- or oligosaccharides. The enzyme has also been shown to have

an inhibitory effect on the growth of yeast and some other pathogens (Bornke et al, 2001). This could indicate that *E. sakazakii* views the ability to metabolize palatinose as advantageous. Knowledge about the α-glucosidase operon could lead to its direct targeting for selective detection, aid in elucidating the pathogen's virulence and potentially lead to its environmental reservoir.

Chapter 3: Research Objectives

The research objectives can be divided into 3 stages of analyses: genotypic, phenotypic, and phylogenetic. The genotypic analysis included determining the prevalence, distribution, and structure of the *pal* operon in *E. sakazakii* and closely related organisms.

The phenotypic analysis used plating on DFI chromogenic agar to determine the phenotype of each strain. Plating on TSA was also done to observe if there was a correlation between the color on DFI and the intensity of color on TSA. M9 Minimal Media plating was also used to determine if the strains can grow when palatinose is the only carbon source. The Biolog Phenotypic Microarray was used to test selective strains and determine the amount of palatinose each uses over a 24 hour period.

The phylogenetic analysis was used to determine if the *pal* operon is under negative selection, if it was horizontally transferred, and if any *pal* genes could have been transferred from *E. rhapontici*. The *pal* gene histories were also compared to eachother to determine each gene's evolutionary history.

The alternative hypothesis of this study was that the various strains of *E. sakazakii* contain most if not all of the *pal* genes with a few aberrant strains missing *pal* genes, since the genes seem to be advantageous to its survival, and are often conserved and present throughout bacterial families. The null hypothesis is the various strains of *E. sakazakii* are missing many of the *pal* genes.

Chapter 4: Materials & Methods

Genotypic Screening with PCR

Primers, seen in Table 4, were designed for PCR to find what *E. sakazakii* strains and related strains in the *Enterobacteriaceae* family were positive for the genes of the α-glucosidase operon. The Bio-Rad DNA Engine Peltier Thermal Cycler was used to conduct PCR with a 55°C annealing temperature for each strain. The PCR cocktail and appropriate concentrations of various reagents used can be seen in Table 5. Gel electrophoresis, using an EmbiTec Electrophoresis Rig at 100V, was done to find which strains were positive for the five *pal* genes studied here. The preliminary results can be seen in Appendix A which is attached. The spreadsheet shows what strains were positive and negative for the 5 α -glucosidase genes screened for this study.

Table 4: *pal* Gene Primers

Gene	Forward Primer	Reverse Primer
palZ	5'-TGG TAT ATC TGG CGC GAC GG-3'	5'-AAA CTG AAA CAC CAT ATT CAG CC-3'
palQ	5'-AAA TGC TCG CCA CCG CGC T-3'	5'-GAT TCA TAA GGT TGC AGT TCC A-3'
palG	5'-GGA TAA CGG CAT CGC CCG CA-3'	5'-ATG AAT TCA CAG GCG CCG A-3'
palK	5'-TCG CCG GGC TTG AGG AGA T-3'	5'-TTC GGC GAG CCG ATA AAC TG-3'
palE	5'-CTG CGC ATG GAG TGC CC-3'	5'-CGT TAC AGG TCA GCC CTT CAT A-3'

Table 5: PCR Cocktail

Amount	Reagent	12 x
5µl	10x PCR buffer (with MgCl) (GeneAmp Applied Biosystems)	60 µl
5 µl	2 mm dNTPs	60 µl
1.3 µl	Forward Primer	16 µl
1.3 µl	Reverse Primer	16 µl
33 µl	Autoclaved Water	396 µl
5 µl	DNA from the many strains extracted using Instagene	60 µl
0.5 µl	TAQ polymerase (ProMega GoTaq DNA Polymerase)	6 µl
		= 50 µl

The *E. sakazakii* strains used for the project were taken from the FDA CFSAN (Center for Food Science and Applied Nutrition) research collection. Each of the strains was confirmed to be *E. sakazakii* by API testing. Many were also submitted by the Nestle Company in Lausanne, Switzerland. Most strains were isolated from infant formula. Strains were also isolated from various milk based powders other than infant formula, fruit powders, toddler cereals and soy cereals. A few other clinical and environmental strains were used along with strains from other various manufacturers. The strains and related *Enterobacter* and other reference strains used is listed in Appendix B.

Sequencing was done using the ExoSAP-IT Protocol (USB) for Amplicon clean up followed by automated dye Sanger dyedoxy sequencing. The ExoSAP reagent was added to the reaction products obtained from PCR (procedure seen in Table 6) to ensure protection of the PCR product, removal of excess dNTPs, and inactivation of residual Taq polymerase. For automated fluorescent sequencing, PCR products were then sent to Amplicon Express (Pullman, WA).

Table 6: Exosap Procedure

ExoSAP Step	Time
Mix 5 µl of post-PCR product with 2 µl Exo-SAP reagent	N/A
Incubate at 37°C	15 minutes
Incubate 80°C	15 minutes
Store at 20°C until sent to Amplicon to be sequences	Indefinitely

Phenotypic Analysis using DFI, TSA & M9

Plating of the strains was done on DFI agar to learn how the different combination of genes affected the phenotypic expression of α -glucosidase activity. DFI agar contains indolyl substrate 5-bromo-4-chloro-3-indolyl-α, D-glucopyranoside. α-glucosidase hydrolyzes the compound, releasing 5-bromo-4-chloro-indolol. This reacts with oxygen to form bromo-chloro-indigo. This is the pigment which creates the blue- green blue- green color of the colony on DFI if the organism possesses α-glucosidase activity (Iverson 2004). Strains were streaked onto the agar using the frozen stocks from the collection and left to incubate for 72 hours. The plates were checked every 24 hours throughout the 72 hour incubation for a positive result.

The strains were also plated onto TSA agar, to see if there was any correlation between the color on TSA and the color on DFI, and therefore if the genotype determines what phenotypic color shows up on both plates. TSA, or Trypticase Soy Agar, contains 17 g tryptone, 3 g soytone, 2.5 g NaCl, 2.5 g K_2HPO_4 and 15 g agar in 1 liter. One of the signifying characteristics of *E. sakazakii* initially was its ability to

form yellow colonies on TSA agar. Therefore, TSA was used as an additional plating step for *E. sakazakii.*

All of the strains were plated on M9 minimal media containing palatinose as the only carbon source. Therefore, unless the strain could metabolize palatinose, the organism would not be able to grow on the media. The recipe for M9 Minimal media can be seen in Table 7.

Table 7: M9 Minimal Media Recipe

M9 Minimal Media Recipe	
M9 salts	Dissolve into 1 L dH_2O
Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g
NH_4Cl	1 g
Add 60 mL to 540 mL molten agar	
1 M $MgSO_4$	Add 0.6 mL to 600 mL of above
0.1 M $CaCl_2$	Add 0.6 mL to 600 mL of above
Vitamin B_1	Add 0.15 mL to 600 mL of above
1% yeast extract	Add 0.6 mL to 600 mL of above
1 M palatinose	Add 12 mL to 600 mL of above

Phenotypic Analysis using Microarray

The Biolog Phenotype Microarray system (Solit 2004) was used to test for metabolic distinctions in the use of palatinose between various strains that have unique combinations of the *pal* genes. That is, PM plates from Biolog will be used to test for time-dependent usage of palatinose. The nutrients and substrates associated with palatinose metabolism were checked for polymorphism between genotypically variable *E. sakazakii* strains. The strains were first plated onto BHI agar and then

were plated on the PM microplates. Well C12 in the biolog plate contains palatinose. An automated densitometric reader measured the nutrient utilization for a specific well in the plate (specific nutrient) over a 24 hours period by ascertaining levels of nitroblue- green tetrazolium oxidized by metabolic activity. If a phenotype is strongly positive in a well, the cells are respiring actively. This reduces the tetrazolium dye and forms a strong color. If the well is weakly positive or negative, it can be interpreted that the respiration is slow or not occurring and no color is formed. In this way, the redox reaction provides for both amplification and quantification of phenotypes. The data are recorded by the Omnilog instrument which captures a digital image of the MicroArray several times each hour and stores the color changes. This can convey if *E. sakazakii* uses the nutrient immediately or at a delayed time.

The results of the phenotypic microarray for each strain collected over a 48 hour period and were then plotted on a kinetic graph. Comparing the kinetic graphs, it can be seen how the genotype of the strain affects the utilization of palatinose over time. A representative sample of the various genotypic combinations of the *pal* genes in *E. sakazakii* were chosen to be tested.

Phylogenetic Analysis

Phylogenetic trees were constructed to show the evolutionary relationships of the genes being studied based on the sequences of the nucleotide homologies of the different alleles. A tree was created for each of the *pal* genes. The tree was compared to a previously constructed strain tree of the housekeeping genes of *E. sakazakii* to check for horizontal transfer of α-glucosidase alleles.

Using these trees, the mean diversity within each clade, or cluster on the tree, was determined. MacClade (Maddison 1994) and PAUP (Swofford 1999) were used to create the trees. The clades were cross referenced with the *pal* genes onto the MLST tree to show shufflings, gains, and losses of the *pal* genes during the evolution of *E. sakazakii*.

Each of the *pal* genes were trees were then reduced for pairwise comparison to eachother. Tanglegrams were created to observe horizontal shuffling of the genes in the various strains. A pairwise ILD (incongruence length difference) test along with a % gene total gene ILD test was done PAUP (Swofford 1999). This test helps to identify the strains that are the recipients of horizontally transferred alleles.

Evolutionary genetic analysis of the *pal* genes substitution patterns was done to determine if the *pal* genes were under negative selection pressure and how they are conserved. The rate of synonymous and nonsynonymous substitution was also determined, as well as the genetic variation of the *pal* genes in comparison to the *pal* genes in *E. rhapontici*. Codon bias was also examined.

The GC content of the genes was determined, which can indicate whether the gene was a recent acquisition to *E. sakazakii* or not. The GC content was then compared to the GC contents of *E. rhapontici* to see if it could have been the organism to transfer the *pal* genes recently.

Chapter 5: Results & Discussion

Genotypic Results

The various strains of *E. sakazakii* showed great variety in the sets of *pal* genes they contained. A total of 31 strains were shown to have all five of the *pal* genes, and another 31 had four out of the five genes, as seen in Figure 2. There were 26 strains that contained three of the genes, 24 strains that contained two of the genes, 16 that only contained one of the genes. Surprisingly, 14 strains did not appear to have any of the genes.

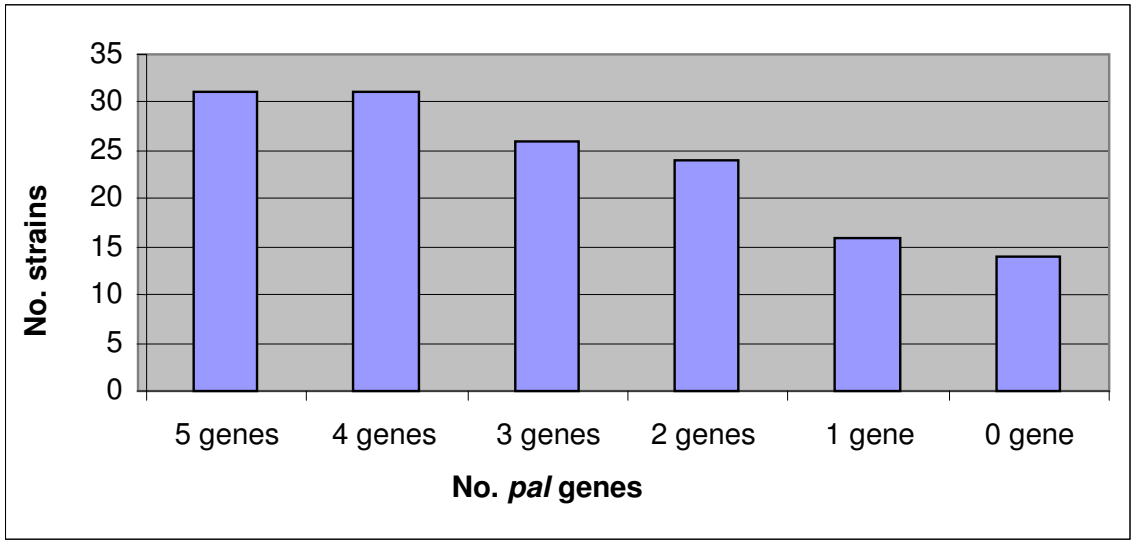


Figure 2: A bar graph showing the number of strains containing each number of genes across the *E. sakazakii* collection.

As seen in Figure 3, each of the genes was present in some combination in about 80 strains, with the exception of *palQ*, which was represented in less then 60 of

the strains. However, all of these strains had palatinose activity. Since *palI*, a gene positioned outside the operon on the *E. sakazakii* chromosome, is homologous to *palQ* in *E. rhapontici*, it is possible that some of the strains contain *palI*, which presumably would convey the same activity that *palQ* does.

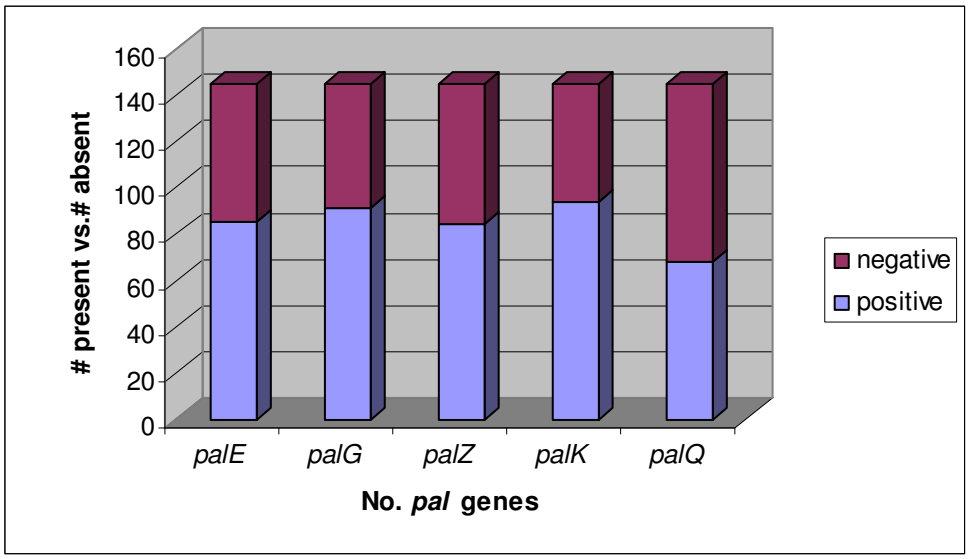


Figure 3: A bar graph showing presence vs. absence of the *pal* genes in *E. sakazakii*. The purple columns show the number of strains that are negative for each of the genes and the blue columns show the number of strains that are positive for each gene.

Phenotypic Results

All of the *E. sakazakii* strains were plated onto M9 Minimal Media, containing palatinose as the only carbon source. Every strain grew on the minimal media. However, not all of the strains had the five *pal* genes that were screened for. These strains may have one or more of the homologous genes that were found in *E.*

rhapontici, such as *palI*, *palR*, or *palF*. *palI* is another α-1,6-glucosidase, *palH* is a α-galactosidase, *palF* is an inner membrane permease, and *palR* codes for a transcriptional regulator (Lehner 2006). It is also possible that there is another component in the *E. sakazakii* genome that conveys palatinose activity but that is unknown at this time. The *E. sakazakii* strains were also plated onto Tryptic Soy Agar (TSA) and DFI agar plates to record the phenotype of the strains and compare the phenotype with the genotype that was obtained from the PCR results. Every strain was plated onto TSA, since yellow colonies on TSA was the long time accepted diagnostic test for the pathogen.

The plates were incubated for 72 hours at 37°C. Every day the color of the colonies was recorded and given a value of +, ++, or +++. On TSA, the + denoted white colonies, the ++ denoted *pal*e yellow colonies, and the +++ denoted bright yellow color. On DFI, + denoted a *pal*e blue- green color with no dot in the middle of the colonies. A ++ denoted a blue- green colony with a small dot in the middle of it. A +++ denoted a vibrant blue- green colony with a dot in the middle that takes up most of the colony or all of it.

Not many of the colonies were +++ at 24 hours, but 68 strains eventually achieved that color after 72 hours. Strains usually reacted the same on both DFI and TSA, either producing a +++, ++, or + on both TSA and DFI consistently. These results can be seen in Figure 4. A table of how each strain reacted on TSA, DFI and M9 minimal media can be seen in Appendix C.

When comparing the phenotypic and genotypic data, several notable a correlations were found between the color displayed on the DFI agar and the *pal*

genes the strain contained. All of the *E. sakazakii* strains that appeared to be a faint blue- green-blue- green- green on DFI after 72 hours, as opposed to the vibrant blue-green blue- green- green color that most have, were missing *palG*. *pal*G encodes for an inner membrane permease. It can be hypothesized that even if the strain has other *pal* genes which create the α-glucosidase enzyme, the enzyme cannot get out of the cell in high enough quantities without an effective permease. Thus, the cell cannot to metabolize enough palatinose and drive the reaction which leads to the vibrant blue-green-blue- green- green color.

It was also found that the strains that lacked all color on DFI lacked *palQ*. *palQ* codes for the enzyme oligo-1,6-glucosidase (the major palatinose degrading enzyme). Therefore, if the glucosidase enzyme is not created, the *E. sakazakii* cannot produce the blue-green color. It may also be hypothesized that these particular strains did not possess *palI*, a secondary glucosidase that lies outside of the immediate 5 gene *pal* region and homologous to a second glucosidase gene found in *E. rhapontici*.

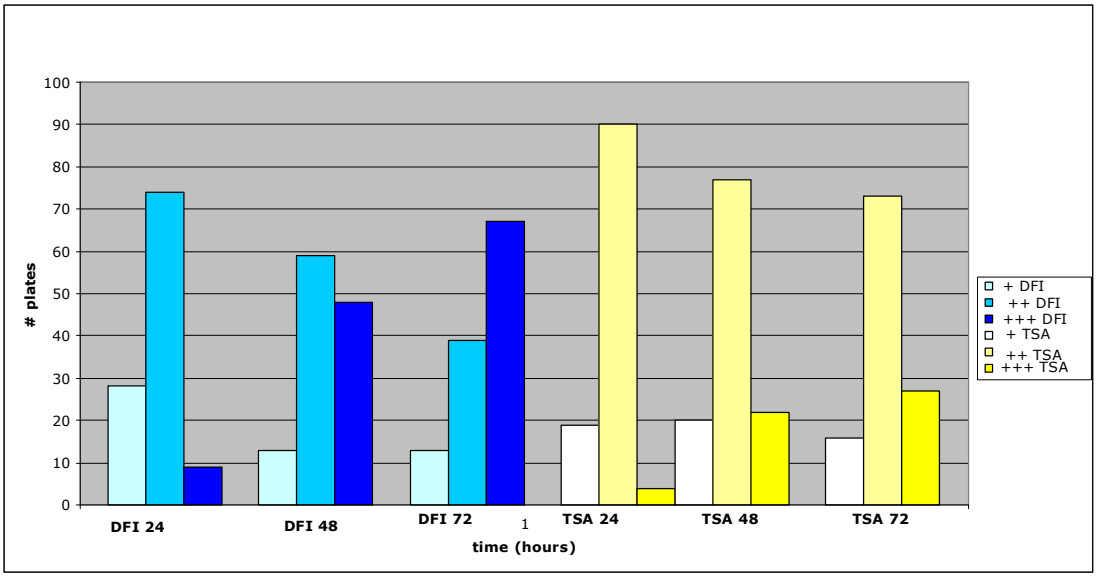


Figure 4: A bar graph showing and comparing the growth of *E. sakazakii* on DFI and TSA agar plates. The color intensity and the number of strains that appeared as such are shown.

The Biolog Phenotypic Microarray (BioLOG Microlog^{TM}) system was used to examine the use of palatinose over a period of time by each of the *E. sakazakii* strains tested. The strains were chosen comprised a representative sample of the various combinations of the *pal* genes to determine how the different combination of *pal* genes affected the catabolism of palatinose over time. Every *E. sakazakii* strain tested appeared to use palatinose at some level and at some point over a 24 hour period. It could be seen that the strains which contained all of the *pal* genes, especially *palG* degraded the most palatinose and did so at the fastest rate. It is notable that in every case there is a delay in the metabolism of palatinose, possibly signifying that palatinose is not the most preferred nutrient among *E. sakazakii* strains.

When the utilization of the palatinose graphs are examined simultaneously with the genes they contain, interesting patterns emerge. For example, strain 1433 contains all of the *pal* genes except *palG*. This strain utilized palatinose but at a much lower and delayed rate when compared to the other strains. Palatinose usage for this strain is barely above baseline. So even though it has the capability of metabolizing palatinose, it is most likely hindered by its ineffectiveness at getting the enzyme out of the cell. Strain 1544 also has the same genotype, but seemed to do better then 1433 at metabolizing palatinose. This could be explained by the possibility of it having the homologous *palF*, also a permease. However, 1544 still did not do as well as 1639, 1680, and 1557 at metabolizing palatinose. Strain 1557 contained all of the *pal* genes and appeared to achieve the highest levels of palatinose metabolism. Strain 1680 did almost as well as 1557. This strain also contains all of the palatinose genes. Strain 1639 contains none of the *pal* genes except *palG*, but appeared to have the ability to metabolize palatinose. This could be because the strain contains some of the other known palatinose metabolizing genes outside of the operon which likely convey the ability to metabolize palatinose. This find also signals the importance of having *palG* to ferry the palatinose enzyme out of the cell. These strains are compared in Figure 5. All of the kinetic graphs done appear in Appendix D. Strains with aberrant genotypic classes of *pal* genes were selected to portray a representative sample of the *E. sakazakii* family.

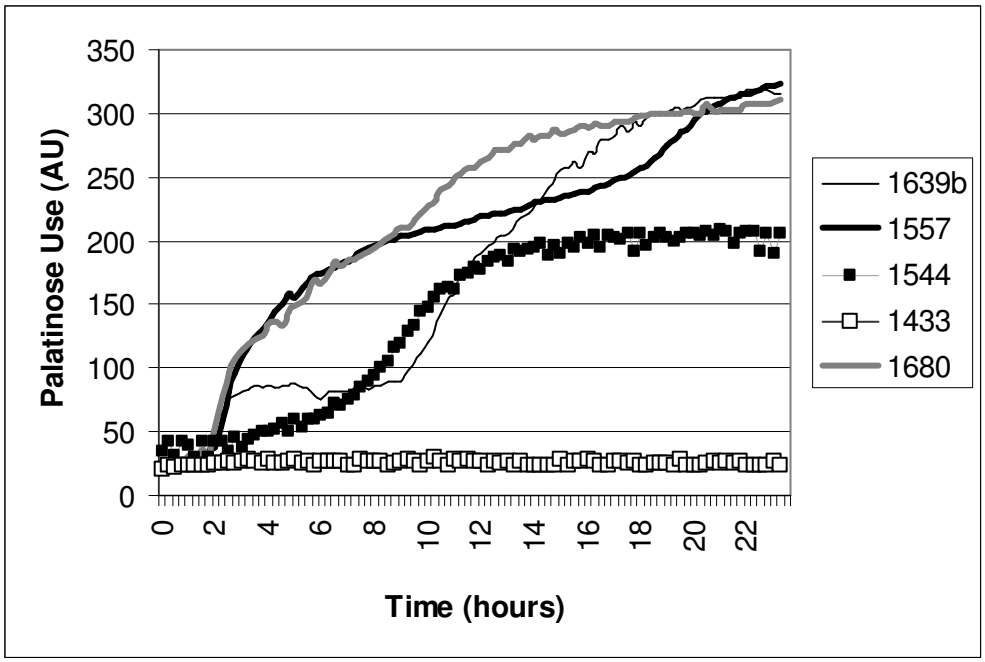


Figure 5: Graph showing a representative sample of 5 strains plotted from the phenotypic microarray data. Strain 1557 contained all 5 *pal* genes and appeared blue-green on DFI. Strain 1433 contains every gene except *palG*, the permease. It showed up as white on DFI. Strain 1544 showed up white on DFI and contains all of the *pal* genes except *palG*. Strain 1639 had none of the *pal* genes except *palG* and showed up white on DFI. Strain 1680 contained all of the *pal* genes and was light blue- green on DFI. AU is an arbitrary unit used by the phenotypic microarray to quantify metabolism by color intensity given off by each strain created by a redox reaction in the well.

Phylogenetic Analysis Results

Obtaining nucleotide sequence data enables the study of gene evolution, since each specific sequence offers much incite into the history of the gene. It is still unknown where *E. sakazakii* received the *pal* genes from. They could all have been received as an operon, through horizontal gene transfer (as opposed to vertical gene transfer), they could have all come in from one organism, such as *E. rhapontici*, at the same time or at different times from different organisms throughout the course of its

evolution. It was also important to examine if the operon is being negatively conserved by the *E. sakazakii* genome and under negative pressure not to mutate. If so, it would indicate that *E. sakazakii* views the *pal* genes as advantageous. Phylogenetic analysis was done to try and answer these questions.

Phylogenetic data was obtained by having the *pal* gene PCR products of each of the strains subjected to nucleotide sequence analysis. The gene sequences were then organized by homologous sequence alignment in ClustalX (Thompson 1997). The sequences were then brought into MEGA 3.0 to do the molecular and evolutionary genetic analysis (Kumar 2001). MEGA compares all of the sequences to eachother and determines the mean overall % divergence, or difference, to eachother.

It was concluded that the *pal* operon is highly conserved by the strains in the *E. sakazakii* family due to the sequence variability seen for all of the codons compared to the sequence variability at the 3rd codon position. The % variability seen for all of the codons was much lower than compared to the % variable at the 3rd codon position. Since 3rd position variations are most often silent changes and evolve at a neutral evolutionary rate, the fact that the other positions have a much lower % variation leads to the conclusion that the operon is under negative pressure to evolve, as seen in Figure 8. Therefore, it is likely that the pathogen sees this operon as advantageous to its survival and as such is maintained under negative selection pressure for function to not chage. *palK* is noteworthy in that it is the one gene that shows to have a much lower 3rd position variation. This could suggest that it was a recent acquisition by *E. sakazakii* compared to the other *pal* genes.

Table 8: Site Variation of *pal* Genes

Locus name	# codons analyzed	% mean variable	% mean 3rd position variability
palZ	120	29.1 %	50 %
palK	142	22.5 %	31.6 %
palG	109	32.1 %	46.7 %
palQ	141	20.1 %	48.9 %
pale	92	21.7 %	42.3%

When allelic variation for each *pal* gene variation was examined for just *E. sakazakii*, the variation was relatively low. When variation between the diversity of the *pal* alleles for each gene in *E. sakazakii* and the *pal* genes of *E. rhapontici* was examined, great variation was seen. Once again, the least variation was seen between *palK* and *E. rhapontici*, suggesting again that *palK* is the one *pal* gene that possibly could have been transferred from *E. rhapontici.* These values can be seen in Figure 9. *palK* also seems to be under less negative selection than the other *pal* genes, since the 3rd position variation in *palK* is very similar to the entire codon variation. Codon bias was also calculated to determine if that could possibly be the cause of the high 3rd position variation and the low genetic variation in the other nucleotides. However, because codon bias is so low, it is not likely attributed to being the cause of low genetic variation. Rather, a more plausible explanation it is attributed to the *pal* genes being under negative selection.

Table 9: Genetic Diversity of *pal* genes in *E. sakazakii* and *E. rhapontici*

Gene	Mean genetic Diversity for *E. sakazakii* only	Mean Genetic Diversity 3rd Position (*E. sakazakii* only)	Average distance of *E. sakazakii* alleles from *E. rhapontici*	Comparison to *E. rhapontici* (3rd position)	Mean Relative Codon Bias
palE	2.9% ± 0.39	6.5%	45.90%	90.25%	2.27
palG	4.9% ± 0.55	10.9%	33.20%	74.5%	2.39
palZ	8.1% ± 0.79	17.1%	42.17%	98.4%	2.20
palK	5.0% ± 0.43	5.7%	28.03%	79.2%	2.27
palQ	3.4% ± 0.40	9.5%	48.6%	95.5%	2.34

The following table, Table 10, shows the evolutionary genetic analysis of the *pal* genes. The d_N value denotes the number of nonsynonymous substitutions that occurred for every 100 codons. The d_S value denotes the number of synonymous substitutions per 100 codons. The M_T value denotes the total mean sequence site variation per 100 codons. The M_3 denotes the mean number of 3rd position variation for every 100 codons. The M_s value denotes the mean number of synonymous substitutions per 100 codons. d_N is the number of nonsynymous substations found. The d_S value is the number of synonymous substitutions, or silent substitutions found. The d_N/d_S value is found by dividing d_N by d_S. Values less than 1 are said to be under negative selection while values over 1 are indicative of genes undergoing positive or dicersifying selection. The closer the d_N/d_S value is to zero the more extensive the

negative selection on the gene. Since all of the genes in the operon have a d_N/d_S value close to zero, it was concluded that the palatinose operon is under substantial negative selection likely as a result of the important functional constraints necessary for palatinose utilization.

Table 10: Evolutionary Genetic Analysis of the *pal* Genes

Locus Name	M_T[1]	M_3[2]	d_S[4]	d_N[5]	d_N/d_S[6]
palZ	7.57 ± 0.68	14.89 ± 1.69	24.15 ± 3.87	4.43 ± 0.762	0.18
palK	3.22 ± 0.33	5.37 ± 0.78	7.12 ± 1.26	2.19 ± 0.33	0.30
palG	4.61 ± 0.483	9.31 ± 1.21	13.27 ± 2.06	2.09 ± 0.35	0.157
palQ	3.31 ± 0.41	8.52 ± 1.09	13.54 ± 2.09	0.781 ± 0.23	0.057
pale	8.43 ± 1.11	6.11 ± 0.97	9.96 ± 1.92	1.05 ± 0.251	0.105

[1] M_T is the mean number of substitutions/100 codons
[2] M_3 is the mean number of 3rd position substitutions/100 codons
[4] dS is the number of synonymous substitutions
[5] dN is the number of nonsynonymous substitutions
[6] dN/dS signifies if the gene is under negative selection

In order to investigate the evolutionary origins of *pal* genes in *E. sakazakii*, the GC content of each of the genes in the palatinose operon was calculated and compared to the GC content of *E. rhapontici pal* homologues to conclude whether or not *E. sakazakii* strains may have received all of some of the operon from this related organism. It was concluded that the GC content was different enough, as seen in

Table 11, to conclude that the genes of the *pal* operon were not acquired recently from *E. rhapontici*. However, *palK* shows the least amount of differences. This could be another sign that *palK* may have been a recent acquisition from *E. rhapontici.* Additionally, these data do not preclude an ancestral acquisition of a *pal* gene from *E. rhapontici* since an old transfer would likely go undetected since ample time would have accrued to ameliorate the codon differences between the various loci.

Table 11: GC Content Comparison

Gene	GC Content % (*E. sakazakii*)	GC Content % (*E. rhapontici*)
palZ	60 %	52 %
palK	53 %	55 %
palG	62 %	54 %
palQ	60 %	55 %
pale	57 %	50 %

Phylogenetic trees were created using WINCLADA (Nixon 1999). In order to further explore the likelihood that this gene family was acquired laterally from horizontal gene transfer into *E. sakazakii* we subjected *pal* gene sequences to the ILD test. The ILD test, or Incongruent Length Difference Test is a method that measures the incongruence between 2 molecular phylogenies. The ILD test evaluates statistically the null hypothesis of congruence between phylogenetic data matrices.

Two matrices are combined together into one larger matrix. Two equally sized submatrices are randomly generated. The tree lengths of the submatrices are compared. If the tree lengths differ, the null hypothesis that the trees are congruent is rejected. The more incongruencies found, the longer the trees are due to the increased branching (Brown 2001).

Table 12 shows the ILD test values for the pairwise comparison of the *pal* genes in *E. sakazakii*. The ILD test, or Incongruence Length Difference Test, is a method that measures the incongruence and evolutionary decoupling of the genes due to horizontal gene transfer for one of both of the genes being compared. It evaluates the null hypothesis the null hypothesis of congruence A value above 0.05 signifies that congruence cannot be rejected and that these genes share a common evolutionary history.

Table 12: ILD Scores for the Pairwise Comparison of *pal* genes

	MLST	*palZ*	*palK*	*palG*	*palQ*	*palE*
palZ	**0.001**	*	**0.001**	**0.002**	**0.001**	0.127
palK	**0.002**	**0.001**	*	0.62	**0.003**	**0.037**
palG	**0.001**	**0.002**	0.62	*	0.07	0.149
palQ	**0.001**	**0.001**	**0.003**	0.07	*	0.59
palE	**0.001**	0.127	**0.037**	0.149	0.59	*

When certain strains were removed that were thought to be the recipients of horizontally transferred *pal* alleles, and the ILD test was conducted again, the ILD value increased to where the congruence of the trees could no longer be rejected. This analysis allowed for the identification of strains that may have received

horizontally transferred *pal* alleles, including *E. sakazakii* the strains 1680, 1681, 1509 and 1672. Overall, this pairwise ILD analysis of each *pal* gene with every other one, suggested that the individual genes themselves are an evolutionary mosaic in *E. sakazakii* such that the majority of the 5 genes that belong to the *pal* operon each have their own evolutionary history decoupled from one another but coming together in a single operon in *E. sakazakii*.

ILD tests were then conducted comparing the *pal* genes with a five gene MLST (Multi Locus Sequence Typing) dataset known to reiterate the evolution and taxonomy of the *E. sakazakii* strains being studied here. This test allowed for an evaluation of whether the *pal* genes we examined have evolved laterally and separate from the *E. sakazakii* strain in which they currently reside. All *pal* genes were found to be highly incongruent with the MLST dataset and had ILD scores of 0.001. To determine which strains may have been horizontally transferred and causing incongruence, the aberrant strains found in the pairwise comparisons were temporarily removed for the ILD test to be run again. For *palQ*, the strains 1681, 1680 and 1563 were taken out of the ILD test. This still resulted in incongruence and an ILD score of less than 0.05. For *palZ*, 1681, 1680, 1563, 1514 and 1673 were taken out of the test. The ILD score was less than 0.05. For *palK*, when strains 1514 and 1673 were taken out, the ILD score was 0.282. For *palE*, 1509 was taken out and the ILD score remained below 0.05. For *palG*, strain 1563 was removed, and again the ILD score remained below 0.05. These data suggested that numerous strains in addition to those listed above were likely recipients of horizontally transferred *pal* alleles. In nearly every case, ILD results implicated several strains as the recipients

of laterally transferred *pal* genes. Furthermore, these findings suggest that *pal* genes are genetically and phylogenetically unstable due to the substantial incidences of horizontal gene transfer in each of these genes' evolution.

Through these phylogenetic tests, horizontal gene transfer of at least a portion of the *pal* genes can be shown. Many strains do not show congruence with the MLST tree of the housekeeping genes of *E. sakazakii*. Since the MLST genes are housekeeping genes which are located throughout the pathogen's chromosome, the incongruence with the *pal* operon shows that it is decoupled from the rest of the chromosome.

The phylogenetic trees created can be seen in figures 17-21 of all of the *pal* genes. The strains that are in a clade, or cluster, together are closely related genetically and evolutionarily.

The pairwise comparison trees confirm that in most cases the genes evolved separately, and not as an entire operon being transferred from one strain to the other. These pairwise comparisons can be seen in figures 23-26. Some of the strains are in separate clades depending on the two *pal* gene trees being compared. The arrows drawn on the pairwise trees show those strains that were identified to cause incongruence using the ILD tests, (displayed as solid black lines), and what strains were identified by sight as being in different incongruent locations between the two trees when looking at the trees side by side (displayed as dashed black lines). Inspection of these trees showed that, like the ILD test, confirmed sporadic congruence. Some strains seemed to be in the same clades, and some strains changed clades between various *pal* genes. Taken together, these findings argue strongly for a

horizontal gene hypothesis to best account for the acquisition and evolution of *pal* genes in *E. sakazakii*. Moreover, these data affirm the genetic and phylogenetic instability of these genes in this pathogenic species further undermining the use of palatinose based enzymatics to serve as diagnostic signatures in *E. sakazakii*. The ILD tests, the pairwise comparisons, the d_N/d_S values calculated, and the comparison of the *pal* gene trees with the MLST trees all lead to the conclusion that at least a portion of the *pal* genes were received by horizontal gene transfer.

Although the environmental reservoir of *E. sakazakii* is unknown, it has been isolated from soil and from cereal based formula. It could be hypothesized that one of its reservoirs is in wheat, which means it would share a host with *E. rhapontici*. This could lead to the conclusion that the transfer of the *pal* genes could occur on wheat between these two closely related species.

Each gene seems to have its own horizontal transfer history and eventually came together to form an operon over time. Even the strains that have all 5 of the *pal* genes may not have received all 5 genes of the operon at once. While some strains may have received all of the genes through vertical gene transfer, some may have received their *pal* genes solely from horizontal gene transfer, and some of the strains could have a mixture of both.

When the distribution of *pal* genes from the original gene screening was mapped as gains and losses onto the *E. sakazakii* MLST strain phylogeny, numerous horizontal introductions of *pal* genes were noted for the 5 different sequences (Fig. 22). Similarly, *palE palZ*, *palK, palG*, and *palQ* were also gained and lost 23, 25, 15, 21, and 26 times, respectively. Together, *pal* genes were either laterally acquired or

laterally lost a total of 110 times for the five genes during the radiation of *E. sakazakii* as a species. These data further confirm a horizontal gene transfer hypothesis for the acquisition of *pal* metabolic function in this pathogen.

The phylogenetic trees of each of the *pal* genes were compared to the MLST tree of the housekeeping genes of *E. sakazakii*. Some of the clades showed shuffling of the strains, indicating that some of the *pal* genes were horizontally transferred and some of the *pal* genes were obtained earlier in *E. sakazakii*'s evolution and they have evolved with the pathogen.

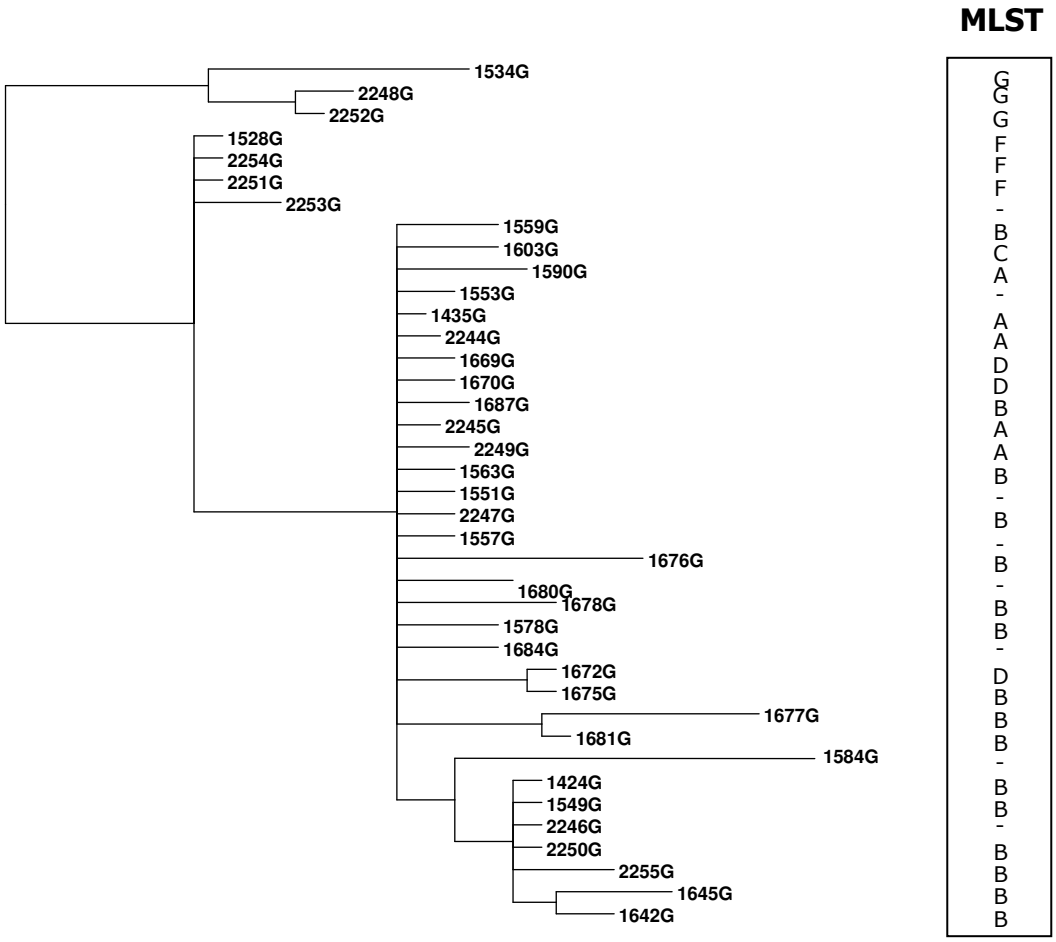


Figure 6: The phylogenetic tree of *palG* in *E. sakazakii.* The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279279)

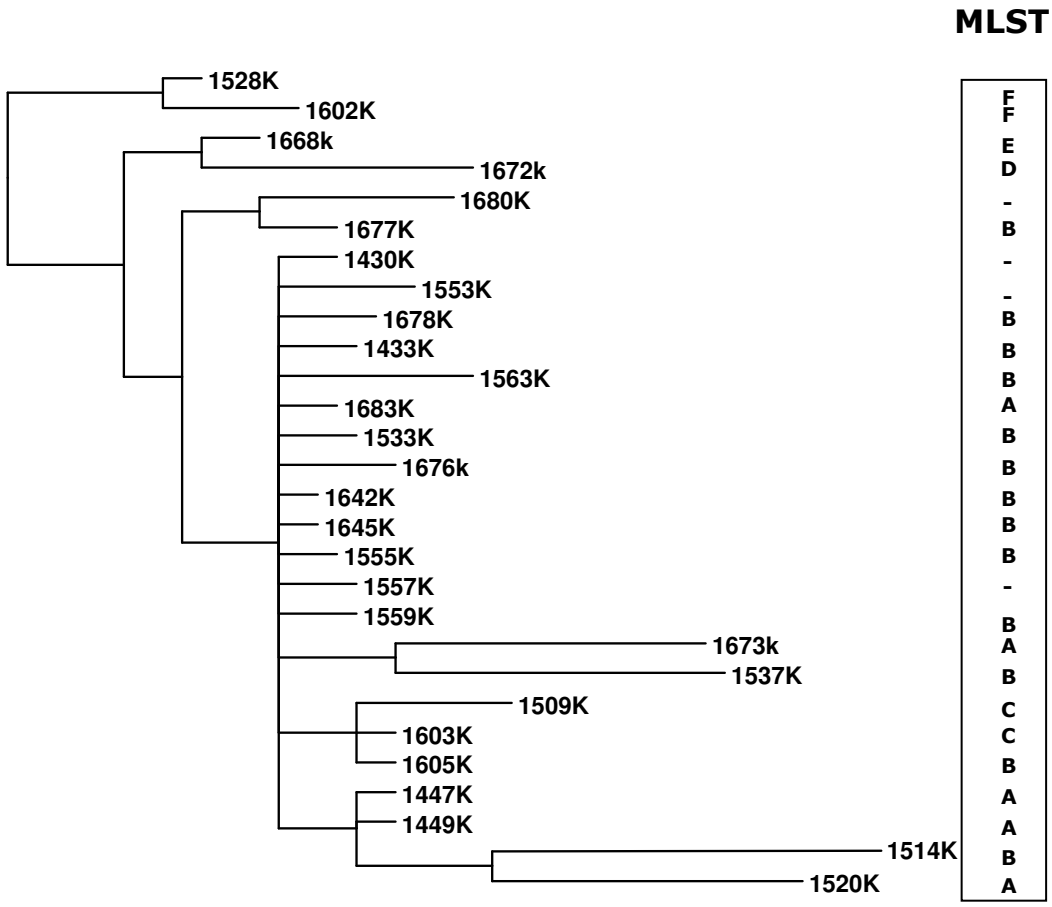


Figure 7: Phylogenetic tree of *palK* gene in *E. sakazakii*. The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279282).

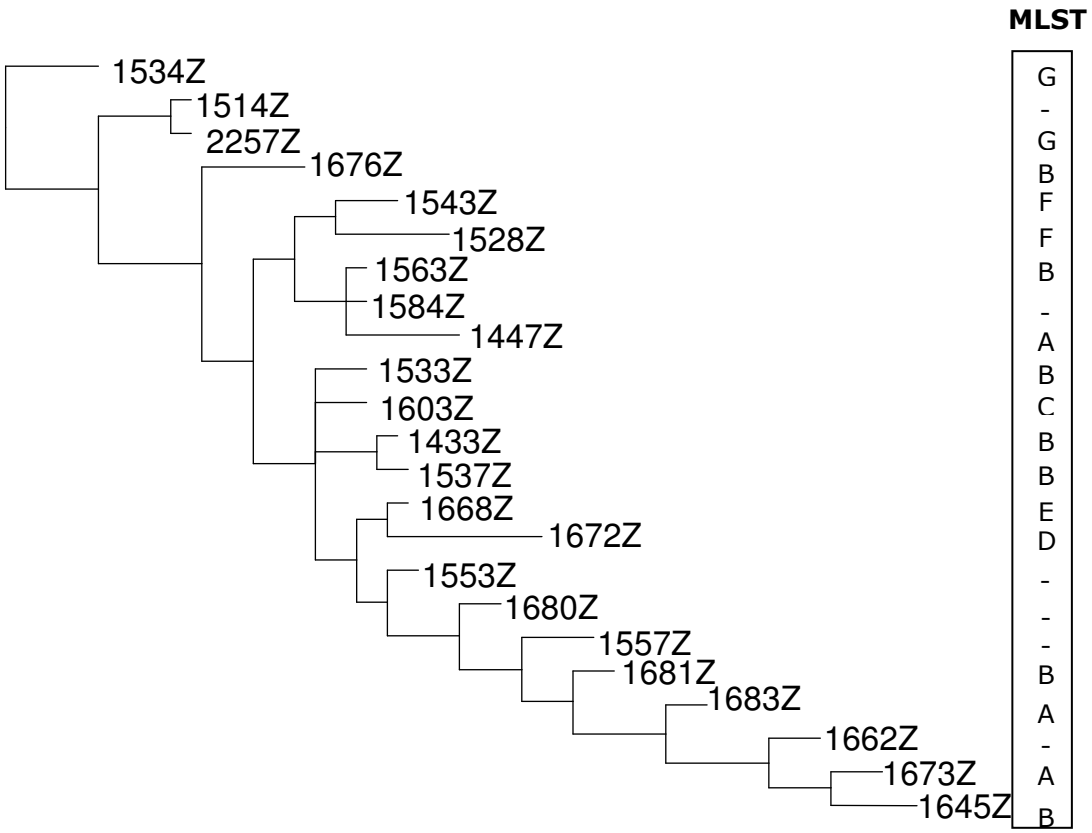


Figure 8: The phylogenetic tree of *palZ* of *E. sakazakii.* The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279285).

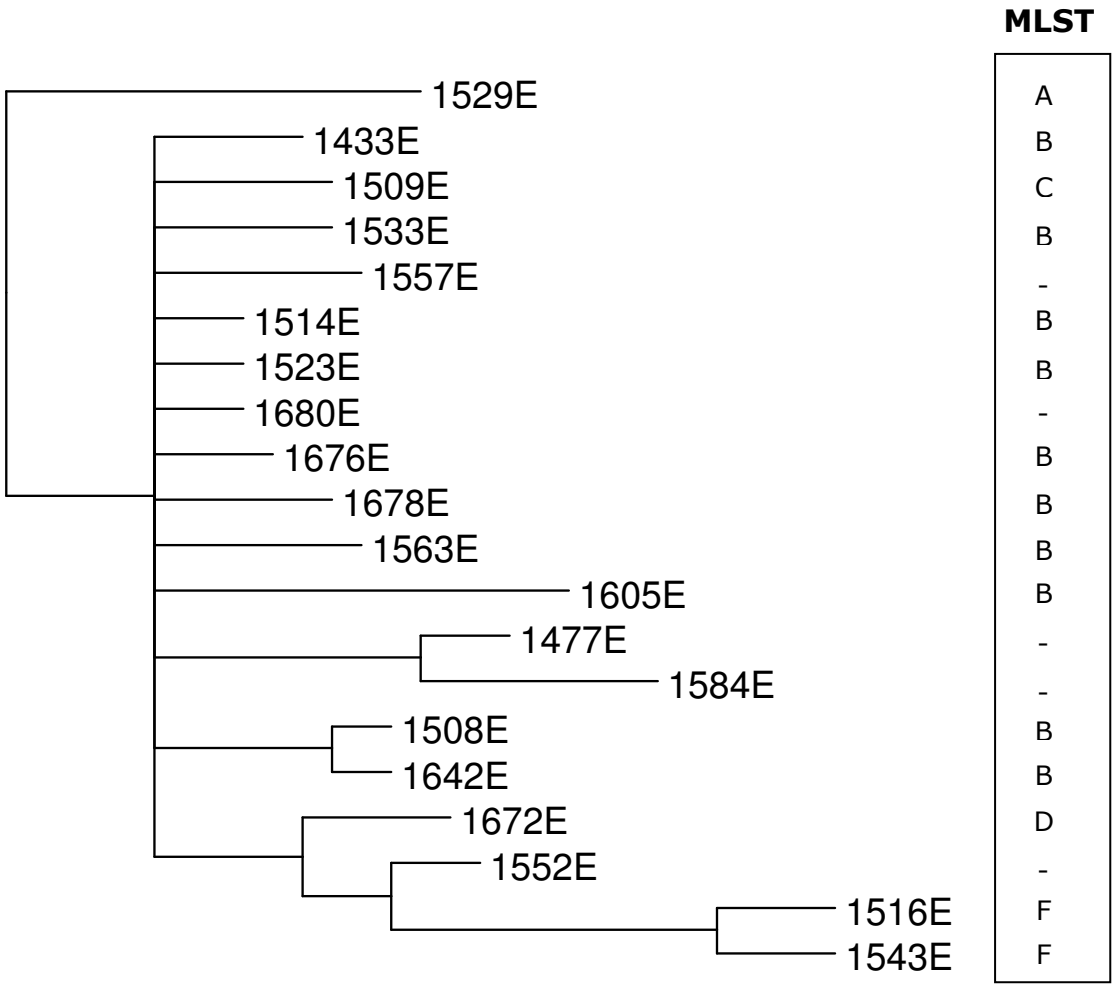


Figure 9: The phylogenetic tree of *palE* in *E. sakazakii*. The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279277).

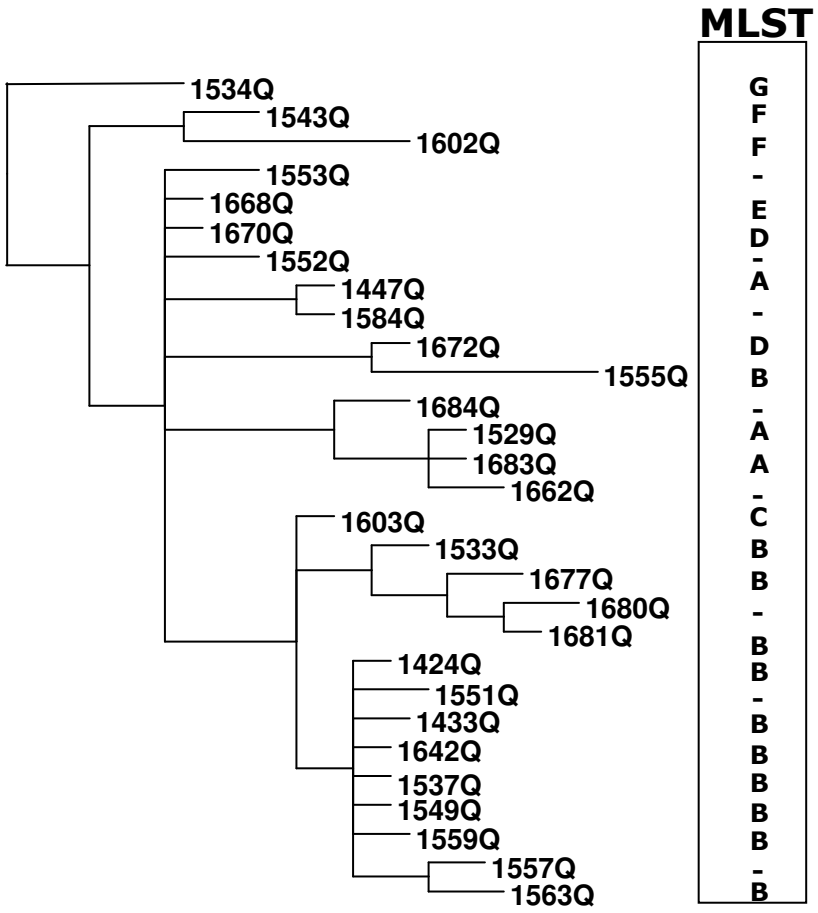


Figure 10: Phylogenetic Tree of *palQ* gene in *E. sakazakii*. The *E. sakazakii* strains that are on the MLST housekeeping gene tree were assigned clade letters. These clade letters are matched to the strains in the *pal* gene tree. The trees were created using WINCLADA (Nixon 1999). The tree was rooted using the homologous gene sequence in *E. rhapontici* (AF279283)

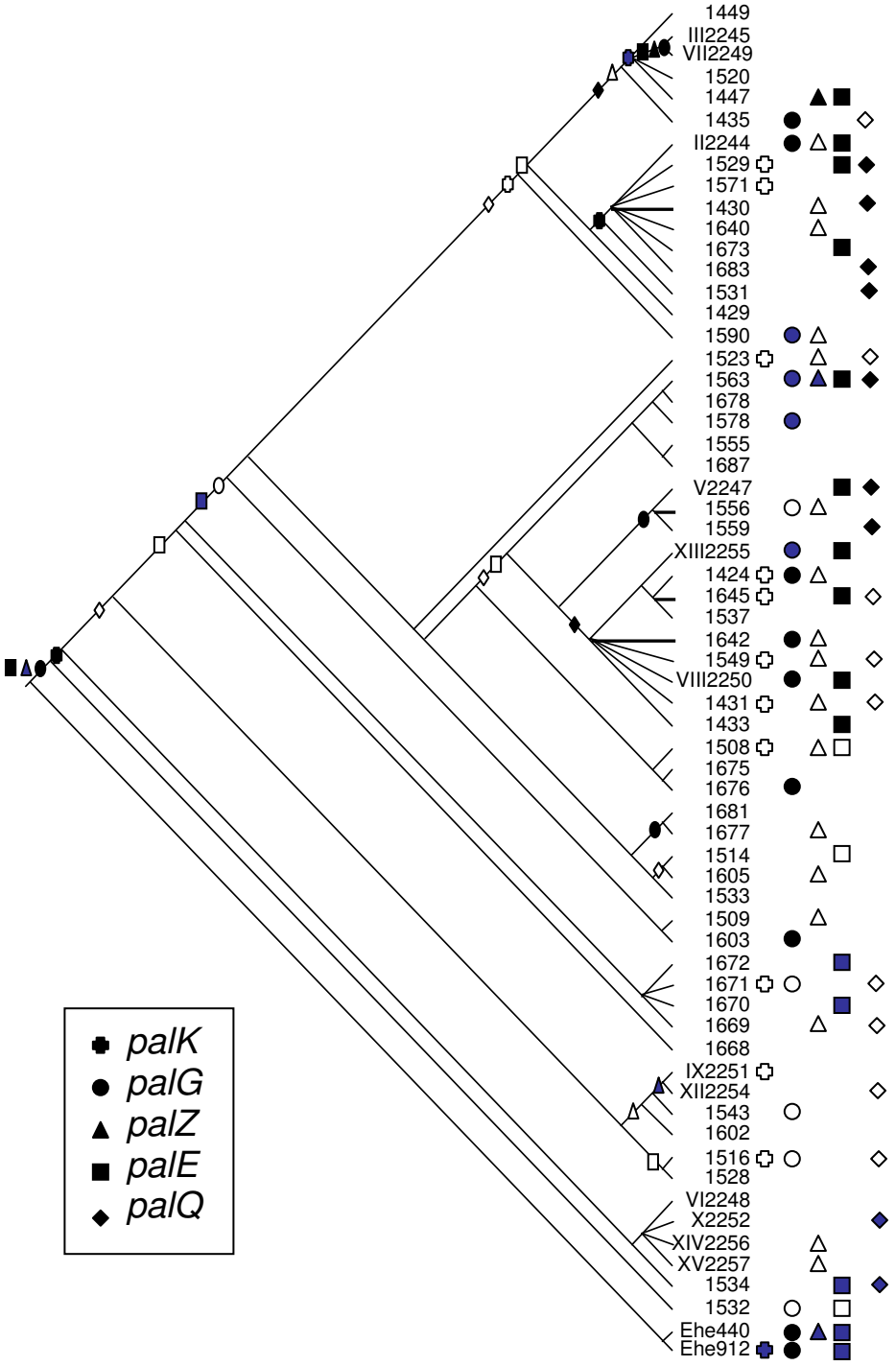


Figure 11: *E. sakazakii* MLST tree of the 5 housekeeping genes. (*icdZ*, *aspC*, *gyrA*, *lysP, clpX*). The MLST tree reflects the overall taxonomy of the *E. sakazakii* strains studied here. The tree shown reflects the minimum number of gains and losses of the 5 *pal* genes during the evolution of *E. sakazakii*.

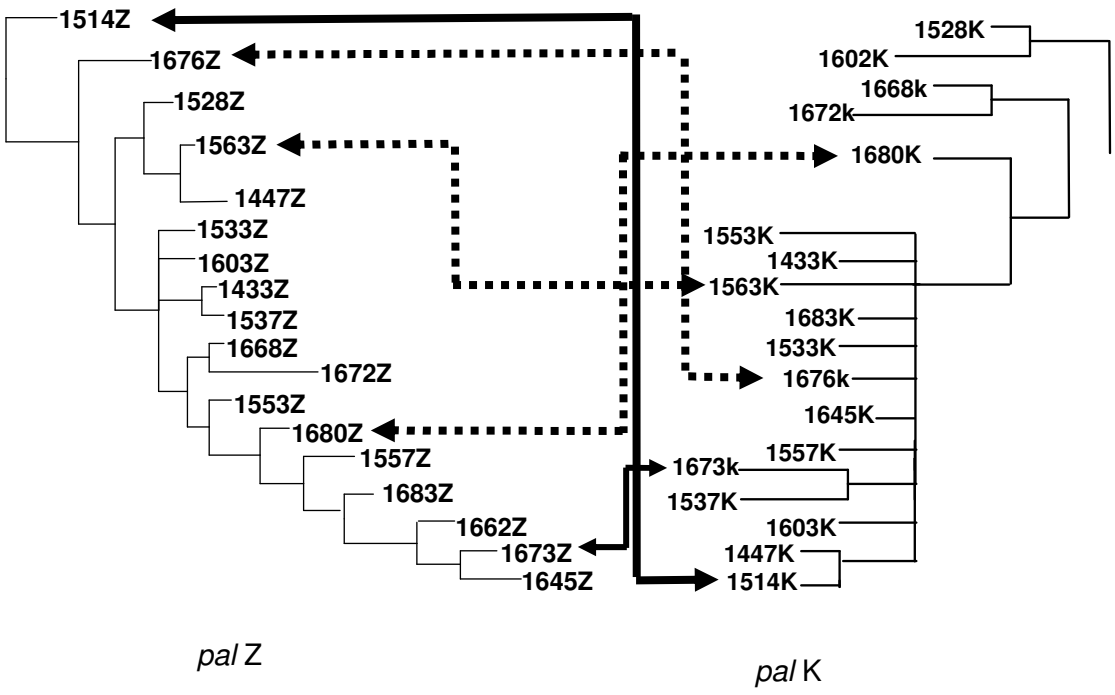


Figure 12: Pairwise comparison tree of *palZ* and *palK*. Strain 1514 and 1673 were identified through ILD tests to be causing incongruence. Strains 1680, 1563, and 1676 were identified by sight as being probable causes of incongruence. It can be seen that they are in different clades in the opposite trees, indicating horizontal gene transfer, while the other strains are located in virtually the same spots on both trees.

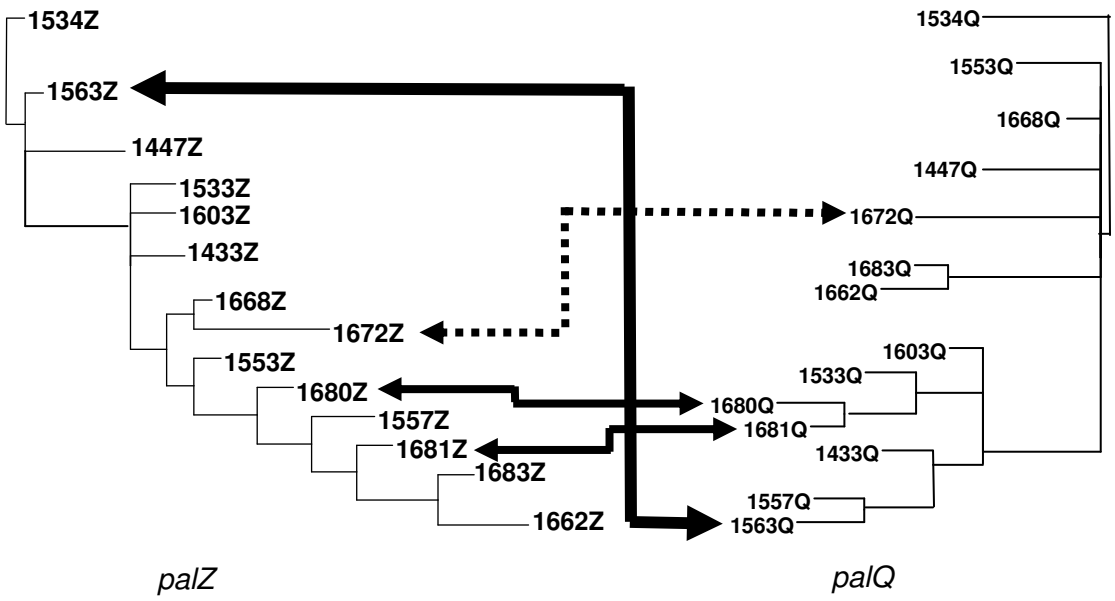


Figure 13: Pairwise comparison of *palZ* and *palQ* trees. Strains 1681, 1680, and 1563 were identified as incongruent by the ILD test. 1672 was identified by sight as being in a different clade in *palZ* as opposed to *palQ*.

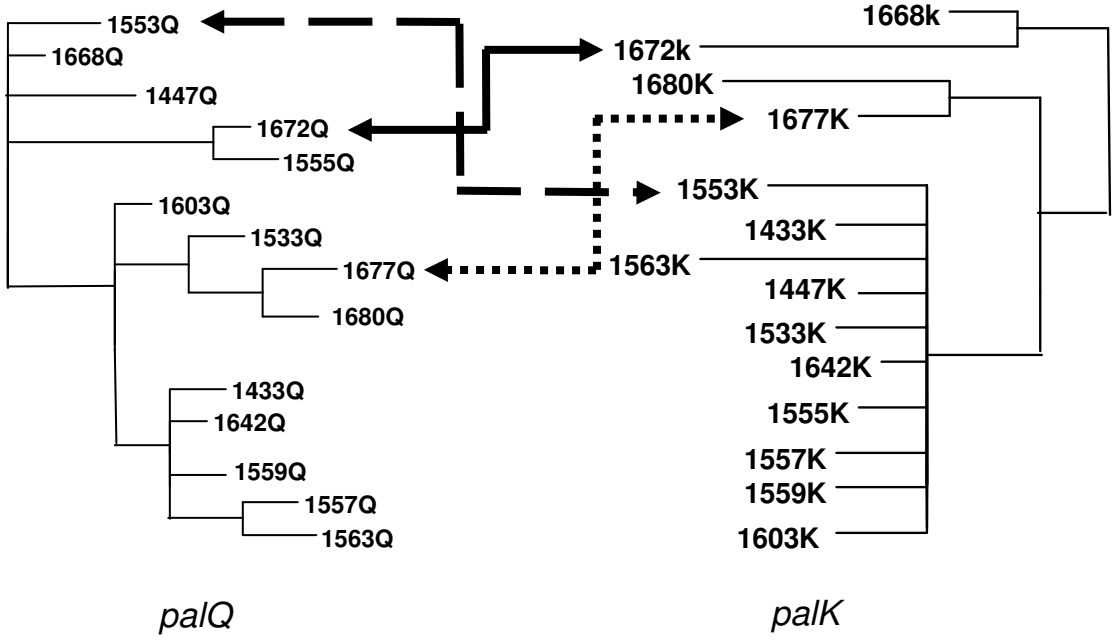


Figure 14: Pairwise comparison of *palQ* and *palK* trees. Strain 1672 was identified by the ILD test. Strains 1677, 1563 and 1553 were identified by sight as causing incongruence.

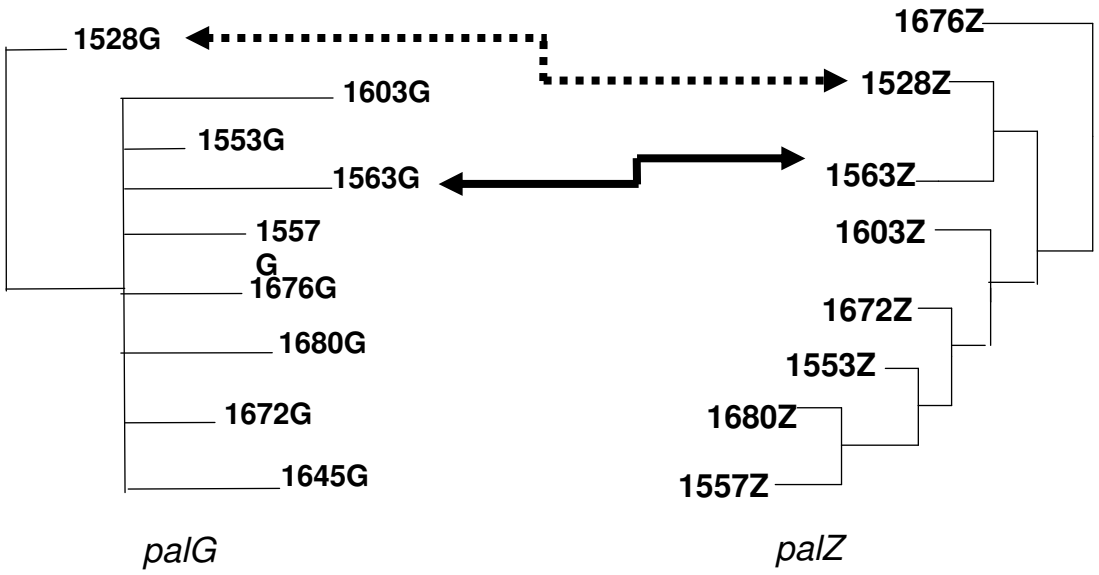


Figure 15: Pairwise comparison of *palG* and *palZ* trees. Strain 1563 was identified as causing incongruence by the ILD test, where as strain 1528 was identified by sight as causing incongruence.

Chapter 6: Conclusions

Several conclusions can be made about the genetic and phenotypic characteristics of *E. sakazakii* from this study. The prevalence of *pal* genes among *E. sakazakii* strains was highly variable. All possible combination of the five *pal* genes examined here in *E. sakazakii* were found, as well as strains containing all or no *pal* genes. It was also notable that several close neighbors of *E. sakazakii* contained several *pal* genes, such as two strains of *E. helvetica*. Given the likelihood of these genes to disseminate horizontally through these strains it is intriguing to postulate expanded α-glucosidase activity among the other species of the *Enterobacteriaceae* family in the future.

The plating experiments yielded substantial insight into the influence of genotype on phenotype among *pal* genes characterized here. Interestingly, *E. sakazakii* strains without *palG* showed a faint blue- green- green color on DFI agar. *palG* is the permease gene which functions in shuttling the mature enzyme out of the cell and into the surrounding media. This indicates how important *palG* is to the *pal* operon and to the metabolism of palatinose such that without *palG* the other *pal* genes present cannot perform to their maximum potential. Due to the observation that all of the strains grew to some extent on M9 minimal media, it is likely that even in the absence of the specific *pal* genes like *pal*G, palatinose metabolism is possible but most likely inefficient. This conclusion was also demonstrated by the strains that were run on the Biolog Phenotypic Microarray.

The few *palG* negative strains that retained substantial ability to metabolize palatinose also may have acquired the homolog of *palF,* a secondary permease gene

found in *Erwinia rhapontici* which was not screed for in this study. We have not tested for this genes presence in *E. sakazakii*, but it is a plausible explanation for function without the presence *palG*.

Other genes on the *E. sakazakii* chromosome with homology to α-glucosidase genes in *E. rhapontici* may also help to explain those *E. sakazakii* strains that had α-glucosidase activity but were negative for the five *pal* genes comprising this operon. For example, *palI* is homologous to *palQ*, the oligo-1,6 glucosidase enzyme, and like *palF* could be present, or could be present in the future as a transfer from *E. rhapontici* in some other part of the *E. sakazakii* chromosome. It was also interesting to note that all of the strains that were not blue- green- green on DFI lacked *palQ*. Once again, all of the strains had enough α-glucosidase activity to be able to grow on M9 minimal media, however they may not have been able to muster the critical concentrations of this enzyme necessary and sufficient for the proper metabolization of palatinose or to create the blue- green color on DFI without *palQ*.

It can be seen through the d_N/d_S values reported here that the α-glucosidase operon is clearly conserved and under substantial negative selection pressure. This is further reinforced in the observed differences between 3rd position codon variation and the general substitution diversity across these genes. The fact that 3rd position substitutions, which are usually silent changes at the amino acid level, greatly outnumber the mean substitutions for the genes overall indicates that the genes are subject to negative selection pressure. *palE* appeared to be one gene that was under more relaxed control by the genome. When the 3rd position variation was calculated per 100 codons, the 3rd position variation was slightly less then the overall variation

per 100 codons for *palE*. However, % 3rd position variation was much greater then overall % variation for *palE*. These findings are interesting in light of the role of *palE* at the surface of the cell where palatinose is recognized and bound prior to digestion.

Horizontal gene transfer was indicated for these genes by several experimental factors. The *pal* operon as a whole was most likely not transferred from *E. rhapontici* due to the varying G-C content and the stark divergence noted between *E. sakazakii pal* alleles and those from *E. rhapontici*. However, *palK* had the closest GC content out of all of the *pal* genes to *E. rhapontici*. Also, its overall % genetic variation was very close to the 3rd position % variation. This indicates that it may have been a more recent acquisition in comparison to the other *pal* genes.

Cladistic analysis of *pal* alleles shows that the several of the strains are in different clades/clusters in comparison to the housekeeping genes representing strain evolution as a whole. If the *pal* genes had avoided lateral transfer and primarily evolved vertically within *E. sakazakii* strains, they would all be in very similar clades when the two trees are compared. Since they are not, it indicates more recent horizontal gene transfer of *pal* genes among and between various *E. sakazakii* strains.

The pairwise comparison of the trees also shows that several of the *pal* genes were not transferred together or at the same time. That is, the pairwise comparison of individual *pal* gene trees shows that several strains are in varying clades on the gene trees, thus some of them must have been transferred into the strains at different times, from different sources, or both.

In conclusion, the genotypic and phenotypic variability of the *pal* operon observed among pathogenic *E. sakazakii* strains makes it a potentially ineffective target for diagnostic test development. Tests primarily focused on such markers could result in false negative results, which could lead to unsafe infant formula, unnecessary illness and product recall. The potential pitfalls of the genomic region in *E. sakazakii* warrant a cautionary note when relying on the degradation of metabolites associated with these genes for detecting this pathogen. Due to increased antibiotic resistance and the presence in infant milk formula, *E. sakazakii* should continue to be monitored to ensure formula safety. These genes, if exploited, can be coupled with additional genetic targets to ensure the complete capture of the pathogen during surveillance.

Appendices

Appendix A Presence Absence Table

set 1	Gene E	Gene G	Gene Z	Gene K	Gene Q	Source
2339	+	+	+	+	+	N/A
2247	+	+	+	+	+	clinical (bronchial wash)
2324	+	+	+	+	+	N/A
1607	+	+	+	+	+	R & F Industries
1669	-	+	-	+	-	R & F Industries
2140	+	-	-	-	+	NA
2235	+	+	+	+	+	NA
2256	+	+	-	+	-	clinical (stool)
1675	+	+	+	+	-	R & F Industries
2255	+	+	+	+	+	environmental (new york)
set 2						
2329	+	+	+	+	-	N/A
2322	+	+	+	+	+	N/A
2327	+	+	+	+	+	N/A
2250	+	+	+	+	+	Clinical
2252	+	+	+	+	+	Clinical
2328	+	+	+	+	-	N/A
2254	+	+	+	+	-	clinical (foot wound)
2326	+	+	+	+	+	N/A
2249	+	+	+	+	+	environmental (new york)
2245	+	+	+	+	+	clinical
set 3						
2330	+	+	+	+	-	N/A
1585	+	+	+	+	-	R & F Industries
1670	+	+	+	+	+	R & F Industries
2258	+	+	+	+	-	clinical (infant)
2325	+	+	+	+	+	N/A
2337	+	+	+	+	+	N/A
2321	+	+	+	+	-	N/A
2331	+	+	-	+	+	N/A
(626)1662	+	+	+	+	-	N/A
set 4						
1584	+	+	+	+	+	R & F Industries
2333	+	+	+	+	+	N/A
2334	+	+	+	+	+	N/A
(614)1639	+	+	-	+	-	N/A
2332	+	+	+	+	+	N/A
2253	+	+	+	+	+	clinical (blood)
1674	+	+	+	+	-	R & F Industries
2251	+	+	+	-	-	clinical (nose)

set 5	Gene E	Gene G	Gene Z	Gene K	Gene Q	Source
1424	-	+	-	-	+	Mammel
1426	+	-	+	-	-	environ-manufacturing plant
1435	-	+	-	-	-	dried infant formula
1445	-	-	-	-	-	Clinical
1584	+	-	+	+	+	R & F Industries
1591	+	-	+	-	-	Nestle Collection
1593	-	-	+	+	+	Nestle Collection
1597	-	-	-	-	-	Nestle Collection
1600	+	-	-	+	-	Nestle Collection
1604	+	+	+	+	+	Nestle Collection
1605	+	-	+	+	-	Nestle Collection
1606	-	-	-	+	-	Nestle Collection
set 6						
1639	-	+	-	-	-	R & F Industries
1642	+	+	-	+	-	R & F Industries
1644	-	-	+	+	-	R & F Industries
2244	+	+	-	+	-	Clinical (stool)
2246	+	+	-	-	-	clinical (sputum)
2248	+	+	+	+	-	clinical (blood)
2313	-	-	+	-	-	N/A
2314	-	+	-	-	-	N/A
2315	+	+	+	+	+	N/A
2316	+	+	+	-	-	N/A
2317	+	+	-	+	-	N/A
set 7						
2318	+	+	-	-	+	N/A
2319	-	-	-	-	+	N/A
2320	-	-	-	+	-	N/A
2323	+	-	-	+	-	N/A
2336	-	-	-	-	+	N/A
2338	-	-	-	-	-	N/A
2340	-	-	-	+	-	N/A
2341	+	-	-	-	-	N/A
613	-	-	-	-	-	N/A
21-1431	-	-	+	+	-	N/A
20-1606	+	-	+	-	-	N/A
2342	+	+	+	+	-	N/A
set 8						
1425	+	-	+	-	+	Mammel
1429	-	-	+	-	-	Clinical
1430	-	-	-	+	+	Canada
1433	+	-	+	+	+	Mammel
1447	+	-	+	+	+	clinical (Israel)
1449	-	-	-	+	+	clinical (Israel)
1508	+	-	-	-	+	environmental – milk
1509	+	-	-	+	+	environmental – milk
1514	+	-	+	+	-	environmental – milk
1516	+	-	+	-	-	environmental – milk
1520	-	-	-	+	+	environmental – milk

set 9	Gene E	Gene G	Gene Z	Gene K	Gene Q	Source
1523	+	-	-	-	-	environmental – milk
1527	-	-	+	+	-	environmental – milk
1528	-	+	+	+	+	environmental – milk
1529	+	-	+	-	+	environmental – milk
1531	-	-	+	+	+	environmental – milk
1532	-	-	+	-	-	environmental
1533	+	-	+	+	+	Environmental
1534	-	+	+	+	+	Environmental
1537	-	-	+	+	+	clinical (infant)
1543	+	-	+	+	+	FDA
1544	-	-	-	-	-	Mead Johnson Plant
set 10						
1549						
1551	-	+	+	-	+	FDA
1552	+	-	-	-	+	FDA
1553	-	+	+	+	+	FDA
1555	-	-	+	+	+	FDA
1556	-	-	-	+	-	FDA
1557	+	+	+	+	+	FDA
1558	-	-	+	+	-	FDA
1559	-	+	+	+	+	FDA
1563	+	+	+	+	+	clinical (infant)
1571	-	-	+	-	-	Comfort Finish Formula
set 11						
1578	-	+	-	+	-	Mead Johnson Plant
1587		-		+	-	Nestle Collection
1590	+	+	-	-	-	Nestle Collection
1602	+	+	-	+	+	Nestle Collection
1603	+	+	+	+	+	Nestle Collection
1640	-	-	-	+	-	R & F Industries
1642	+	+	-	+	+	R & F Industries
1645	+	+	+	+	-	R & F Industries
1662	-	+	+	+	+	R & F Industries
1667	+	-	-	-	-	R & F Industries
1668	-	+	+	+	+	R & F Industries
1671	-	-	+	-	-	R & F Industries
set 12						
1672	+	+	+	+	+	R & F Industries
1673	+	-	+	+	-	R & F Industries
1676	+	+	+	+	-	R & F Industries
1677	+	+	-	+	+	R & F Industries
1678	+	+	-	+	-	R & F Industries
1680	+	+	+	+	+	R & F Industries
1681	+ faint	+	+	+	+	R & F Industries
1683	-	-	+	+	+	R & F Industries
1684	-	+	-	-	+	R & F Industries
1687		+	+	+	-	R & F Industries
1893	-	-	-	+	-	R & F Industries

Set 13	Gene E	Gene G	Gene Z	Gene K	Gene Q	Source
EH 440	+	+	+	+	-	Nestles
EH 912	+	+	-	-	-	Nestles
PA 2134	+	-	-		-	Nestles
Ki 2135		-	-	+	-	Nestles
Cf 2136		-	-	-	-	Nestles
Ea 2137	+	+	-	-	-	Nestles
Eac 2138		-	-	-	-	Nestles
Cas 2139		-	-		-	Nestles
Ec 2141	-	-	+	-	-	Nestles
set 14						
Ec 2142		+	-	+	-	Nestles
Eg 2143	-	-	-	-	-	Nestles
Eh 2144	-		-	-	-	Nestles
Ck 2145	-	-	-	-	-	Nestles
Ed 2146			-	-	-	Nestles
Eca 2147	-		-	-	-	Nestles
Bxv 2257	+	+	-	+	-	Nestles
ET 23730	-		-	+	+	Nestles

Appendix B List of all of the strains used in this project.

	Strain		Strain		strain
E. sakazakii	613	*E. sakazakii*	1602	*E. sakazakii*	2316
E. sakazakii	1424	*E. sakazakii*	1603	*E. sakazakii*	2317
E. sakazakii	1425	*E. sakazakii*	1604	*E. sakazakii*	2318
E. sakazakii	1426	*E. sakazakii*	1605	*E. sakazakii*	2319
E. sakazakii	1429	*E. sakazakii*	1606	*E. sakazakii*	2320
E. sakazakii	1430	*E. sakazakii*	1607	*E. sakazakii*	2321
E. sakazakii	1433	*E. sakazakii*	1639	*E. sakazakii*	2322
E. sakazakii	1435	*E. sakazakii*	1640	*E. sakazakii*	2323
E. sakazakii	1445	*E. sakazakii*	1642	*E. sakazakii*	2324
E. sakazakii	1447	*E. sakazakii*	1642	*E. sakazakii*	2325
E. sakazakii	1449	*E. sakazakii*	1644	*E. sakazakii*	2326
E. sakazakii	1508	*E. sakazakii*	1645	*E. sakazakii*	2327
E. sakazakii	1509	*E. sakazakii*	1662	*E. sakazakii*	2328
E. sakazakii	1514	*E. sakazakii*	1667	*E. sakazakii*	2329
E. sakazakii	1516	*E. sakazakii*	1668	*E. sakazakii*	2330
E. sakazakii	1520	*E. sakazakii*	1669	*E. sakazakii*	2331
E. sakazakii	1523	*E. sakazakii*	1670	*E. sakazakii*	2332
E. sakazakii	1527	*E. sakazakii*	1671	*E. sakazakii*	2333
E. sakazakii	1528	*E. sakazakii*	1672	*E. sakazakii*	2334
E. sakazakii	1529	*E. sakazakii*	1673	*E. sakazakii*	2336
E. sakazakii	1531	*E. sakazakii*	1674	*E. sakazakii*	2337
E. sakazakii	1532	*E. sakazakii*	1675	*E. sakazakii*	2338
E. sakazakii	1533	*E. sakazakii*	1676	*E. sakazakii*	2339
E. sakazakii	1534	*E. sakazakii*	1677	*E. sakazakii*	2340
E. sakazakii	1537	*E. sakazakii*	1678	*E. sakazakii*	2341
E. sakazakii	1543	*E. sakazakii*	1680	*E. sakazakii*	2342
E. sakazakii	1544	*E. sakazakii*	1681	*E. sakazakii*	(614)1639
E. sakazakii	1549	*E. sakazakii*	1683	*E. sakazakii*	(626)1662
E. sakazakii	1551	*E. sakazakii*	1684	*E. sakazakii*	20-1606
E. sakazakii	1552	*E. sakazakii*	1687	*E. sakazakii*	21-1431
E. sakazakii	1553	*E. sakazakii*	1893		
E. sakazakii	1555	*E. sakazakii*	2235	*Citrobacter freundii*	Cf 2136
E. sakazakii	1556	*E. sakazakii*	2244	*Citrobacter koseri*	Ck 2145
E. sakazakii	1557	*E. sakazakii*	2245	*E. asburiae*	Eas 2139
E. sakazakii	1558	*E. sakazakii*	2246	*E. dissolvens*	Ed 2146
E. sakazakii	1559	*E. sakazakii*	2247	*E. gergoviae*	Eg 2143
E. sakazakii	1563	*E. sakazakii*	2248	*E. helvetica*	EH 440
E. sakazakii	1571	*E. sakazakii*	2249	*E. helvetica*	Eh 2144
E. sakazakii	1578	*E. sakazakii*	2250	*E.Helvetica*	EH 912
E. sakazakii	1584	*E. sakazakii*	2251	*Enterobacter amnigenus*	Ea 2137
E. sakazakii	1584	*E. sakazakii*	2252	*Enterobacter cancerogenous*	Eac 2138
E. sakazakii	1585	*E. sakazakii*	2253	*Enterobacter cancerogenous*	Eca 2147
E. sakazakii	1587	*E. sakazakii*	2254	*Enterobacter cloaceae*	Ec 2141
E. sakazakii	1590	*E. sakazakii*	2255	*Enterobacter cloaceae*	Ec 2142

E. sakazakii	**1591**	***E. sakazakii***	**2256**	***Enterobacter cloaceae***	**Ec 2140**
E. sakazakii	**1593**	***E. sakazakii***	**2258**	***K. intermedia***	**Ki 2135**
E. sakazakii	**1597**	***E. sakazakii***	**2313**	***P. agglomerans***	**PA 2134**
E. sakazakii	**1600**	***E. sakazakii***	**2314**		**Bxv 2257**
		E. sakazakii	**2315**		**ET 23730**

Appendix C: Table showing the DFI, TSA and M9 data.

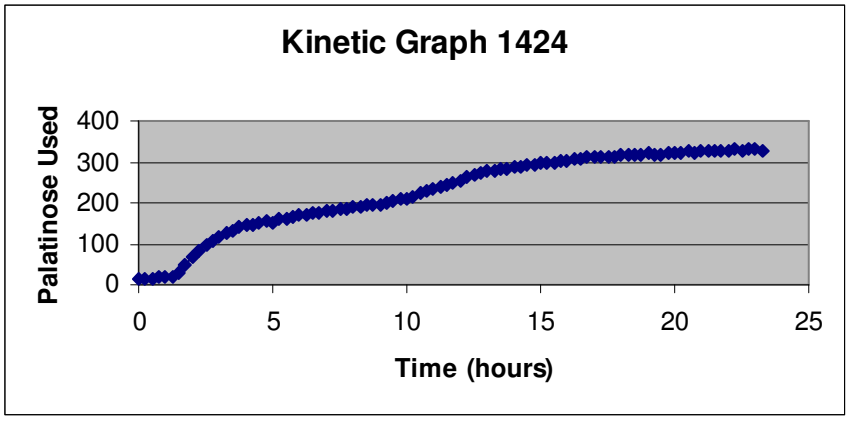
	DFI			TSA			M9
strain	24 hours	48 hours	72 hours	24 hours	48 hours	72 hours	
613	++	++	++	++	++	++	+++
1424	+	+	+	++	+++	+++	+++
1425	+++	+++	+++	+++	+++	+++	+++
1426	++	++	+++	++	++		+++
1429	+++	+++	+++	++	++	++	+++
1430		+++	+++	++	++	++	+++
1433	+	+	+	+	+	+	+++
1435	++	+++	+++	++	++		+++
1445	+	+	+	++	++	++	+++
1447	++	++	++	++	++	++	+++
1449				++	++	++	+++
1508	+	+	+	++	++	++	+++
1509	+	+	+	++	++	++	+++
1514	++	++	++	++	++	++	+++
1516	++	++	++	+	+	+	+++
1520	++	+++	+++	+++	+++	+++	+++
1523	++	+++	+++	+++	+++	+++	+++
1527	++	++	++	+	++	++	+++
1528	++	+++	+++	++	++	++	+++
1529	++	++	++	++	++	++	+++
1531	++	++	+++	++	++	++	+++
1532	+	+	+	+	+	+	+++
1533	+++	+++	+++	+	+	+	+++
1534	++	+++	+++	++	++	++	+++
1537	++	+++	+++	++	+++	+++	+++
1543	+	++	++	+	+	++	+++
1544	-	-	-	-	-	-	+++
1549	++	+++	+++	-	-	-	+++
1551	++	++	++	++	++	++	+++
1552	++	+++	+++	++	++	++	+++
1553	++	+++	+++	++	++	++	+++
1555	++	++	++	++	++	++	+++
1556	++	++	++	++	++	++	+++
1557	++	++	+++	++	++	++	+++
1558	++	++	++	++	++	++	+++
1559	++	++	++	++	++	++	+++
1563	++	++	+++	++	++	++	+++
1571	++	++	+++	+	+	+	+++
1578	++	++	+++	++	++	++	+++
1584	++	++	++	++	++	++	+++
1585	++	++	++	-	+	++	+++
1587	++	++	++	+	++	++	+++
1590	++	+++	+++	++	++		+++

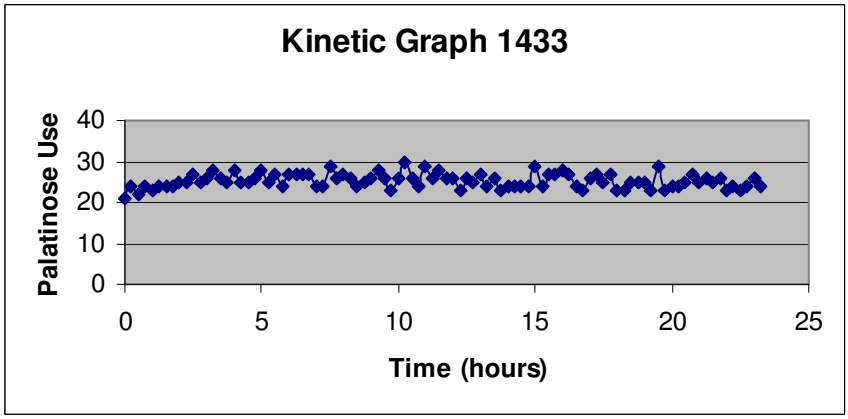
1591	++	+++	+++	-	++	+++	+++
1593	+	++	++	++	++	++	+++
1597	++	++	++	++	+++	+++	+++
1600	+	+	+	-	+	+	+++
1602	++	++	+++	+	+	+	+++
1603	+++	+++	+++	++	++	++	+++
1604	+++	+++	+++	++	+++	+++	+++
1605	++	++	+++	++	++	++	+++
1606	++	+++	+++	++	+++	+++	+++
1607	++	++	+++	+	+	+	+++
1639	-	-	-	++	++	++	+++
1640	+++	+++	+++	+++	+++	+++	+++
1642	+	++	+++	+	+	+	+++
1644	++	+++	+++	++	++	+++	+++
1645	+	++	++	++	++	++	+++
1662	++	++	++	++	++	++	+++
1667	-	-	-	+	+	+	+++
1668	++	+++	+++	++	+++	++	+++
1669	++	+++	+++	++	++	++	+++
1670	+	++	++	++	+++	+++	+++
1671	++	+++	+++	++	++	++	+++
1672	++	++	++	++	+++	+++	+++
1673	++	+++	+++	++	++	++	+++
1674	+	++	++	++	++	++	+++
1675	++	+++	+++	++	++	++	+++
1676	++	++	+++	++	+++	+++	+++
1677	++	++	+++	++	++	++	+++
1678	++	+++	+++	++	++	++	+++
1680	++	++	++	++	++	++	+++
1681	++	++	++	++	++	+++	+++
1683	++	++	+++	++	++	++	+++
1684	++	++	++	++	++	++	+++
1687	++	+++	+++	++	++	++	+++
1893	++	+++	+++	++	+++	+++	+++
2235	+	+	+	++	++	++	+++
2244	-	-	-	-	-	-	+++
2245	+	+	+	-	-	-	+++
2246	+++	+++	+++	++	++	++	+++
2247	+	+	+	+	+	+	+++
2248	++	+++	+++	++	++	++	+++
2249	-	+	+	-	-	-	+++
2250	+	+++	+++	+++	+++	+++	+++
2251	+	++	++	+	+	+	+++
2252	++	+++	+++	++	++	++	+++
2253	++	++	+++	+	+	+	+++
2254	-	++	++	Pink	pink	pink	+++
2255	-	++	++	-	-	-	+++
2256	-	-	-	-	-	-	+++

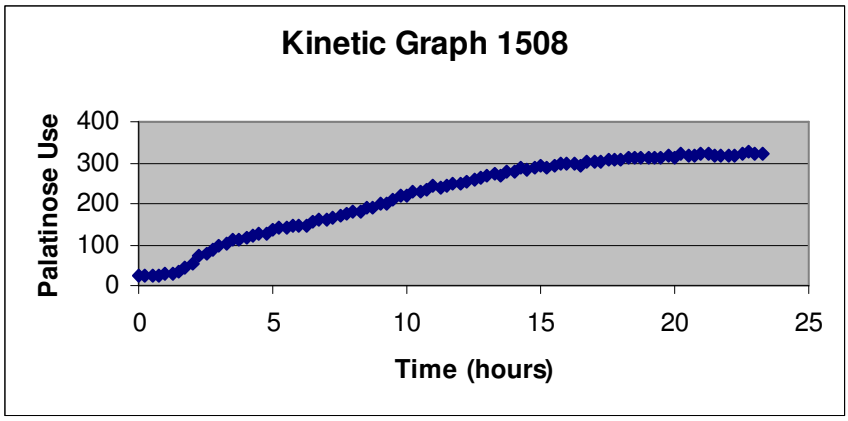
2258	+	+	++	++	++	++	+++
2313	++	++	++	++	++	++	+++
2314	++	++	+++	++	++	++	+++
2315	++	+++	+++	+	++	++	+++
2316	++	++	+++	++	++	++	+++
2317	+	++	++	++	+++	+++	+++
2318	+	++	++	++	++	++	+++
2319	++	+++	+++	+	++	+++	+++
2320	+	+	+	++	++	++	+++
2321	++	+++	+++	-	+	++	+++
2322	++	++	+++	++	++	++	+++
2323	++	++	+++	++	+++	+++	+++
2324	++	+++	+++	++	++	++	+++
2325	+	++	+++	++	++	++	+++
2326	+	++	++	++	+++	+++	+++
2327	++	++	++	++	++	++	+++
2328	++	+++	+++	++	++	++	+++
2329	+	+++	+++	++	++	++	+++
2330	+	++	++	++	++	+++	+++
2331	+	++	++	++	++	+++	+++
2332	++	+++	+++	++	+++	+++	+++
2333	++	+++	+++	++	+++	+++	+++
2334	++	+++	+++	-	++	++	+++
2336	+	++	++	++	++	++	+++
2337	++	+++	+++	++	+++	+++	+++
2338	+++	+++	+++	++	++	++	+++
2339	++	+++	+++	++	+++	+++	+++
2340	+++	+++	+++	++	++	++	+++
2341	++	+++	+++	++	++	++	+++
2342	++	+++	+++	++	++	++	+++
(614) 1639	-	-	-	-	-	-	+++
(626) 1662	-	-	-	-	+	+	+++
20-1606	++	++	+++	++	+++	+++	+++
21-1431	++	++	++	++	++	+++	+++
Cf 2136	black	black	black				
Ck 2145	-	-	-	-	-	-	
Eas 2139	++	++	++	-	-	-	
Ed 2146	-	-	-	-	-	-	
eg 2143	-	-	-	-	-	-	
eh 440							
eh 2144	-	-	-	+	+	+	
eh 912							
ea 2137	-	-	-	-	-	-	
eac 2138	-	-	-	++	++	++	

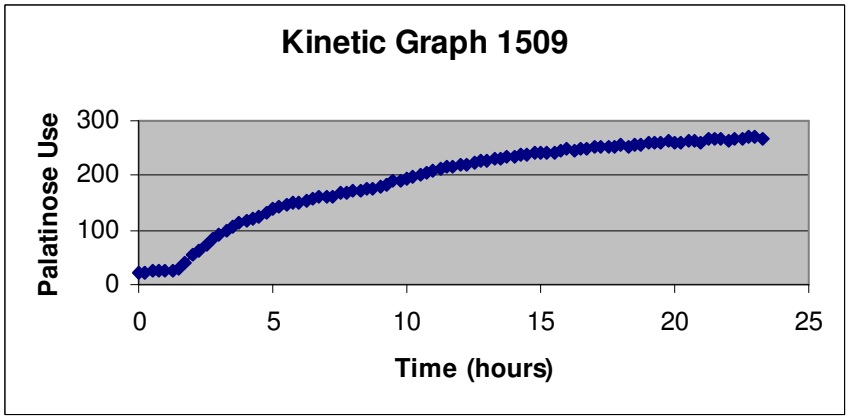
eca 2147	-	-	-	+	+	+	
ec 2141	-	-	-	+	+	+	
ec 2142	-	-	-	+	+	+	
ec 2140	-	-	-	-	-	-	
ki 2135	++	++	++	-	-	-	
pa 2134	-	-	-	++	++	++	
bxv 2257	++	++	++	++	++	++	
et 23730							
1453	++	++	++	++	++	++	
614	-	-	-				

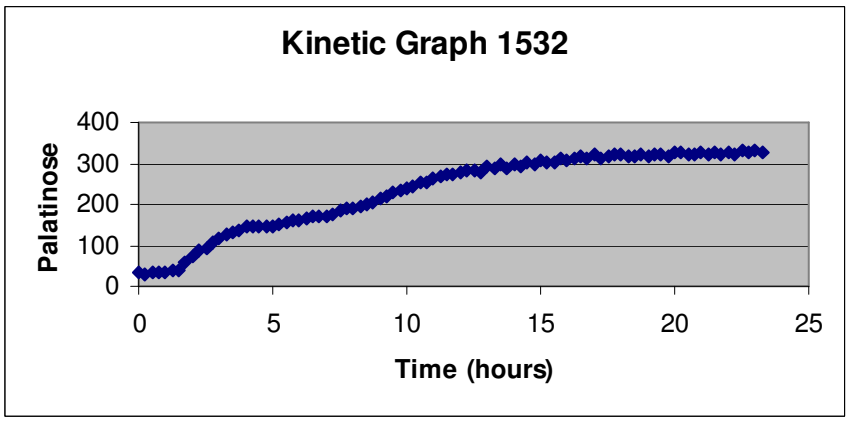
Appendix D: Phenotypic Microarray Kinetic Graphs

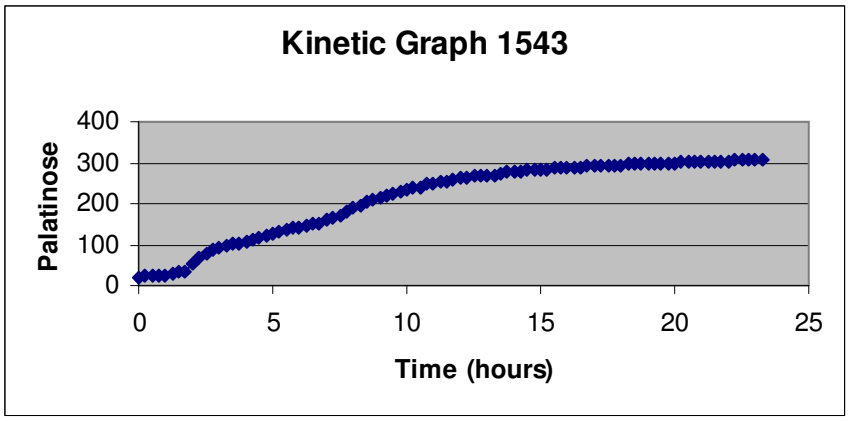


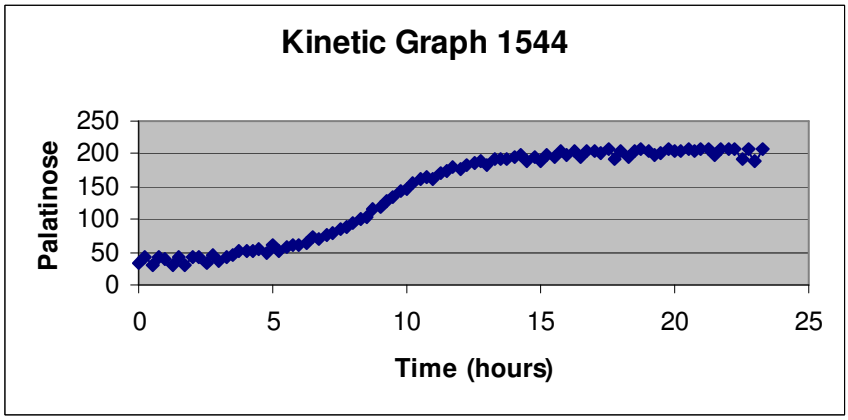


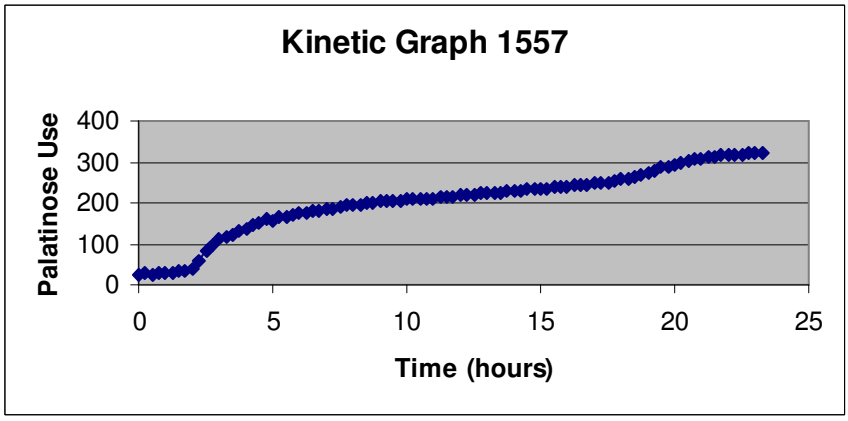

Kinetic Graph 1508
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

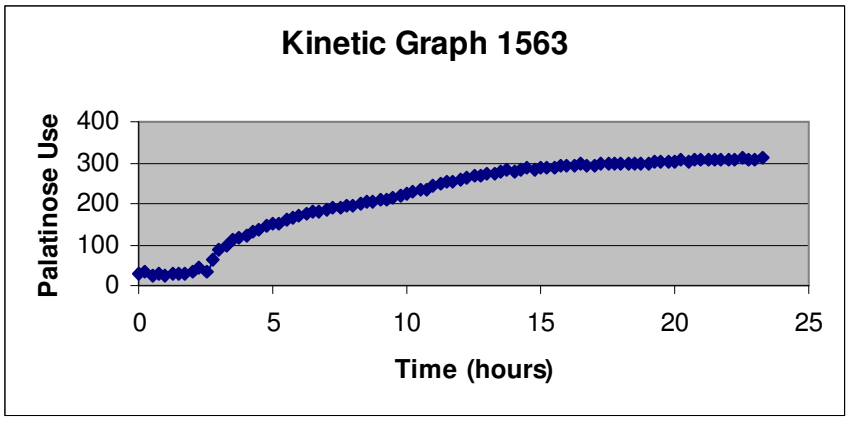

Kinetic Graph 1509
Palatinose Use
300
200
100
0
0
5
10
15
20
25
Time (hours)

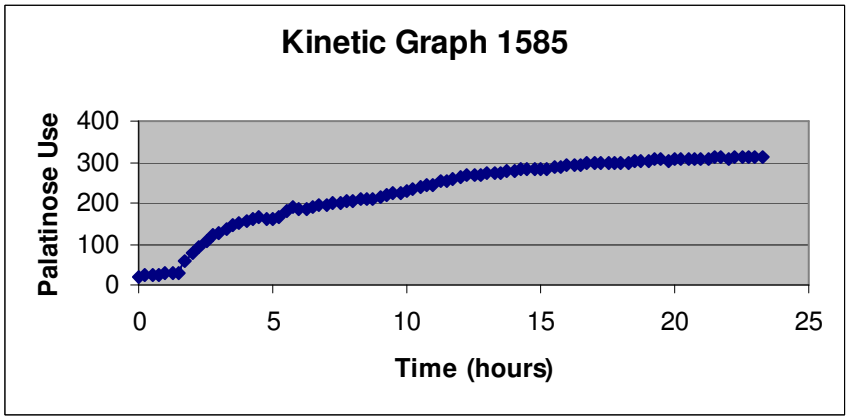

Kinetic Graph 1532
Palatinose
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

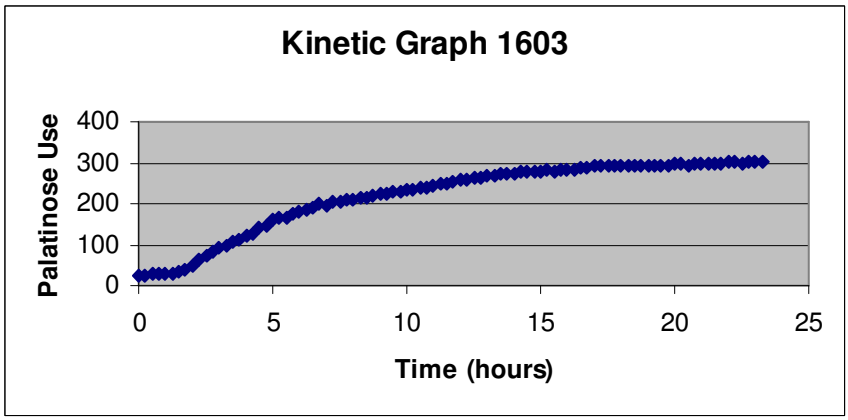

Kinetic Graph 1543
Palatinose
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

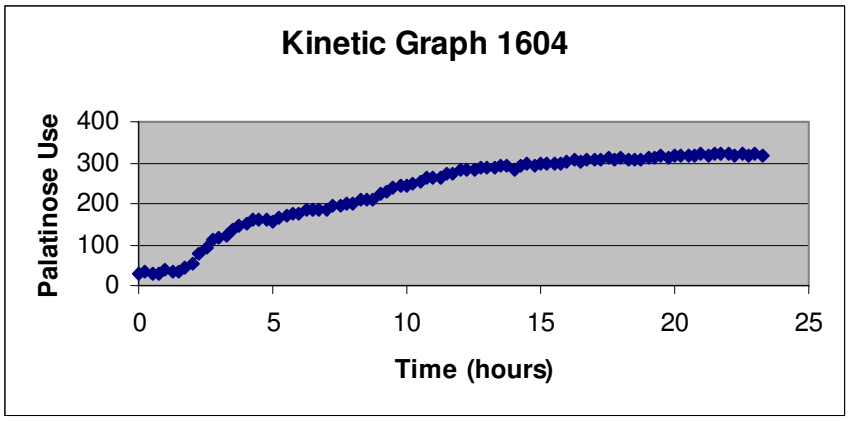

Kinetic Graph 1544
Palatinose
250
200
150
100
50
0
0
5
10
15
20
25
Time (hours)

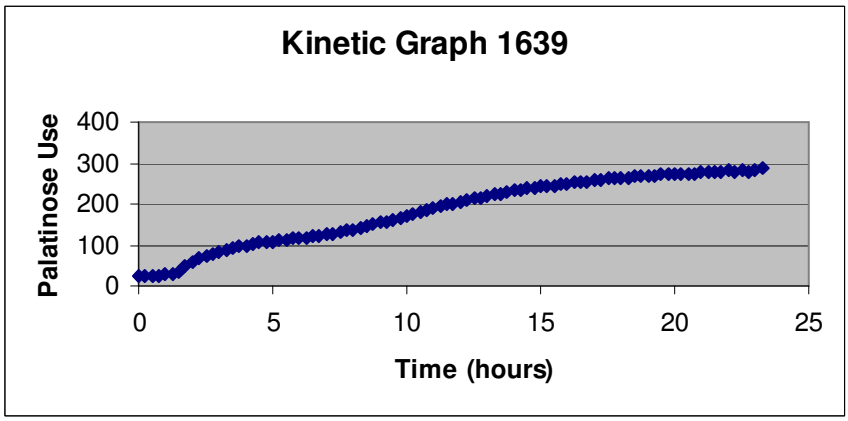

Kinetic Graph 1557
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

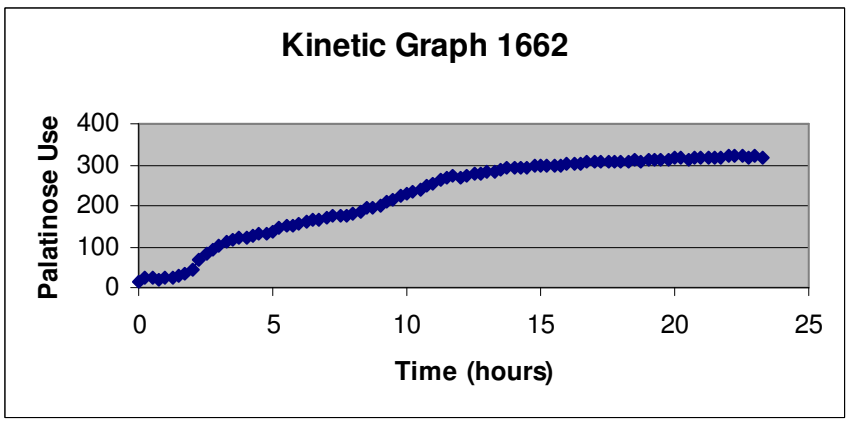

Kinetic Graph 1563
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

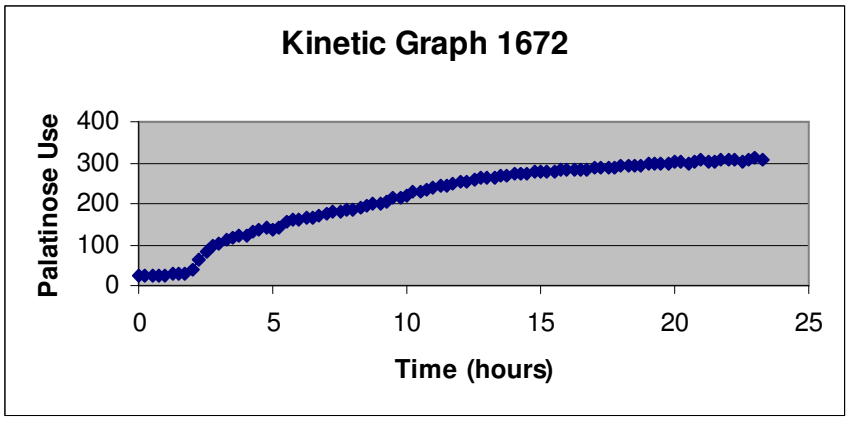

Kinetic Graph 1585
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

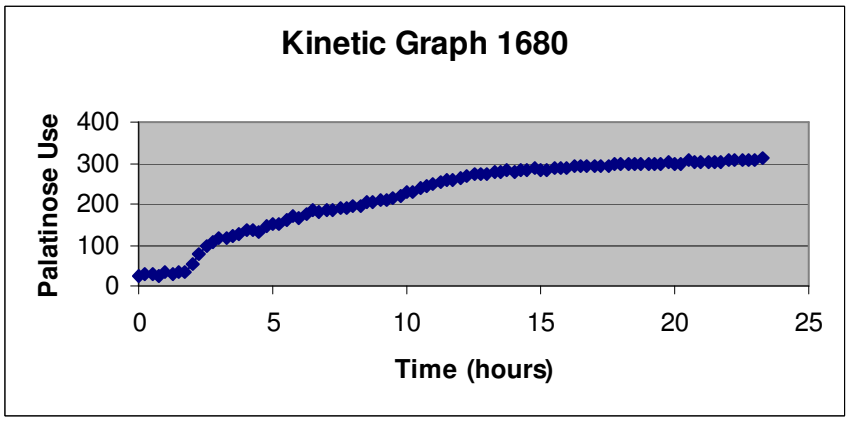

Kinetic Graph 1603
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)


Kinetic Graph 1604
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)


Kinetic Graph 1639
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)


Kinetic Graph 1662
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)


Kinetic Graph 1672
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)


Kinetic Graph 1680
Palatinose Use
400
300
200
100
0
0
5
10
15
20
25
Time (hours)

Glossary of Abbreviations

DFI – Droggan-Forsythe-Iverson, a chromogenic media in which *E. sakazakii* colonies appear blue green on and *Salmonella* colonies appear black

DNA – deoxyribonucleic acid

dN – number of nonsynonymous substitutions in the gene

dS – number of synonymous substitutions in the gene

GC content – the percentatage of the gene's genetic sequence that is made up of guanine and cytosine, as opposed to adenine and thymine

HGT – horizontal gene transfer; the transfer of genes from other bacterial organisms

ILD – incongruence length difference test; conducted to determine congruence and incongruence amongst gene trees

M_3 – 3rd position variation per 100 codons

M_S – synonymous variation per 100 codons

M_T – nonsynonymous variation per 100 codons

M9 minimal media – media which contains palatinose as its only carbon source

MEGA – Molecular Evolutionary Analysis Software

PAUP – Phylogenetic Analysis Using Parsimony; Computer program that uses the principles of maximum parsimony to reconstruct the phylogenetic relationships of bacterial strains or genes.

palE – gene which codes for a sugar binding protein

palG – gene which codes for an inner membrane permease

palK – gene which codes for an ATP binding cassette protein

palQ – gene which codes for the glucosidase enzyme

palZ – gene which codes for the glucosidase enzyme

PCR – polymerase chain reaction; method to make multiple copies of genes using Primers

TSA – Tryptone Soy Agar

VGT – vertical gene transfer; the transfer of genes from parent cells

Bibliography

Asakura H, Morita-Ishihara T, Yamamoto S, Igimi S. 2007. Genetic characterization of thermal tolerance in *Enterobacter sakazakii*. Microbiology and Immunology. 51:671-677.

Bowen AB, Braden CR. 2006. Invasive *Enterobacter sakazakii* disease in infants. Emerging Infectious Diseases. 12: 1185-1189.

Bornke F. Hajirezael M, Sonnewald U. 2001. Cloning and characterization of the gene cluster for palatinose metabolism from the phytopathogenic bacterium *Erwinia rhapontici*. Journal of Bacteriology. 183: 2425 – 2430.

Brown EW, LeClerc JE, Kotewicz ML, Cebula TA. 2001. Three R's of bacterial evolution: How replication, repair, and recombination frame the origin of species. Environmental and Molecular Mutagenesis 38:248-260.

Corti G, Panunzi I, Losco M, Buzzi R. 2007. Postsurgical osteomyelitis caused by *Enterobacter sakazakii* in a healthy young man. Journal of Chemotherapy. 19: 94-96.

Friedemann M. 2007. *Enterobacter sakazakii* in food and beverages (other than infant formula and milk powder). International Journal of Food Microbiology. 116: 1-10.

Grimm M, Stephan R, Iverson C, Manzardo GG, Rattei T, Riedel K, Ruepp A, Frishman D, Lehner A. 2008. Cellulose as an extracellular matrix component present in *Enterobacter sakazakii* biofilms. Journal of Food Protection. 71:13-18.

Gurtler JB, Kornacki JL, Beuchat LR. 2005. *Enterobacter sakazakii*: A coliform of increased concern to infant health. International Journal of Food Microbiology. 104:1-34.

Gurtler JB, Beuchat LR. 2007. Growth of *Enterobacter sakazakii* in reconstituted infant formula as affected by composition and temperature. Journal of Food Protection. 70: 2095-2103.

Derzelle S, Dilasser F, Maladen V, Soudrie N, Leclercq A, Lombard B, Lafarge V. 2007. Comparison of three chromogenic media and evaluation of two molecular-based identification systems for the detection of *Enterobacter sakazakii* from environmental samples from infant formulae factories. Journal of Food Protection. 70: 1678-1684.

Iverson C, Druggan P, Forsythe S. 2004. A selective differential medium for

Enterobacter sakazakii, a preliminary study. International Journal of Food Microbiology. 96: 133-139.

Iverson C, Lehner A, Mullane N, Bidlas E, Cleenwerck I, Marugg J, Fanning S, Stephan R, Joosten H. 2007. The taxonomy of *Enterobacter sakazakii*: proposal of a new genus Cronobacter gen. nov. and descriptions of Cronobacter sakazakii comb. Nov. Cronobacter sakazakii subsp. Sakazakii comb. Nov., Cronobacter sakazakii subsp. Malonaticus subsp. Nov., Cronobacter turicensis sp. nov., Cronobacter muytjensii sp. nov., Cronobacter dublinensis sp. nov. and Cronobacter genomospecies 1. BMC Evolutionary Biology. 17:64.

Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. MEGA2: Molecular Evolutionary Genetic Analysis Software. Bioinformatics (in press).

Lin LC, Beuchat LR. 2007. Survival of *Enterobacter sakazakii* in infant cereal as affected by composition, water activity, and temperature. Food Microbiology. 24: 767-777.

Lin LC, Beuchat LR. 2007. Survival and growth of *Enterobacter sakazakii* in infant cereal as affected by composition, reconstitution liquid and storage temperature. Journal of Food Protection. 70: 1410-1422.

Lehner A, Riedel K, Rattei T, Ruepp A, Frishman D, Breeuwer P, Diep B, Eberl L, Stephan R. 2006. Molecular characterization of the α-glucosidase activity in *Enterobacter sakazakii* reveals the presence of a putative gene cluster for palatinose metabolism. Systematic and Applied Microbiology. 29: 609 – 625.

Maddison WP, Maddison DR. 1994. MacClade v. 3.04 program and documentation. Sinauer Associates, Sunderland MA.

Mohan Nair MK, Venkitanarayanan K. 2007. Role of bacterial OmpA and host cytoskeleton in the invasion of human intestinal epithelial cells by *Enterobacter sakazakii*. Pediatric Research. 62: 664-669.

Mullane NR, Iverson C, Healy B, Walsh C, Whyte P, Wall PG, Quinn T, Fanning S. 2007. *Enterobacter sakazakii* an emerging bacterial pathogen with implications for infant health. Minerva Pediatrica. 59: 137-148.

Mullane NR, Murray J, Drudy D, Prentice N, Whyte P, Wall PG, Parton A, Fanning S. 2006. Detection of *Enterobacter sakazakii* in dried infant milk formula by cationic-magnetic-bead capture. Applied and Environmental Microbiology. 72: 6325 – 6330.

Mullane NR, Whyte P, Wall PG, Quinn T, Fanning S. 2007. Application of pulsed-field gel electrophoresis to characterize and trace the prevalence of

Enterobacter sakazakii in an infant formula processing facility. International Journal of Food Microbiology. 116: 73-81.

Nazarowec-White M, Farber JM. 1999. Phenotypic and genotypic typing of food and clinical isolates of *Enterobacter sakazakii*. Epidemiological Typing. 48:559-567.

Nixon KC. 1999. WINCLADA v.0.9.9b. program and documentation. Cornell University, New York.

Pei X, Guo Y, Xu J, Liu X. 2007. Antibiotic susceptibility of *Enterobacter sakazakii* isolated from powdered formula for infant and young children. Journal of Hygiene Research. 36: 63-65.

Raghav M, Aggarwal PK. 2007. Purification and characterization of Enterobacter sakazakii enterotoxin. Canadian Journal of Microbiology. 53: 750-755.

Reidel K, Lehner A. 2007. Identification of proteins involved in osmotic stress response in *Enterobacter sakazakii* by proteomics. Proteomics. 7: 1217 – 1231.

See KC, Than HA, Tang T. 2007. *Enterobacter sakazakii* bacteraemia with multiple splenic abscesses in a 75 year old woman: a case report. Age Aging. 36: 595-596.

Seo KH, Bracket RE. 2005. Rapid, specific detection of *Enterobacter sakazakii* in infant formula using a real-rime PCR Assay. Journal of Food Protection. 68: 59-63.

Solit R. BioLOG Microlog ™ system, Release 4.2 User Guide. © 2004, 2003, 2001. BioLOG Inc. All Rights Reserved.

Shaker R, Osali T, Al-Omary W, Jaradat Z, Al-Zuby M. 2006. Isolation of *Enterobacter sakazakii* and other *Enterobacter* sp. from food and food production environments. Food Control. 18: 1241 – 1245.

Swofford DL. 1999. Phylogenetic analysis using parsimony (PAUP v. 4.03b) program and documentation. The Smithsonian Institution, Washington D.C.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins Dg. 1997. The ClustalX Windows Interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acid Resistance 25: 4876 – 4882.

Torres-Chavolla E, Ramirez-Cerda E, Gutierrez-Rojo R. 2007. Isolation and identification of *Enterobacter sakazakii* in infant formula. Foodborne

Pathogens and Disease. 4:164 – 168.

Townsend S, Barron JC, Loc-Carillo C, Forsythe S. 2006. The presence of endotoxin in powdered infant formula milk and the influence of endotoxin and *Enterobacter sakazakii* on bacterial translocation in the infant rat. Food Microbiology. 24: 67-74.

1677K